Monte Carlo study of *trans*- and *cis*-azobenzene bulk phases with coarse-grained two-site Gay–Berne model

Master's thesis

Perttu Tuovinen

Degree programme in physics
Faculty of science
University of Oulu

2019

Contents

Acknowledgements

This thesis arose out of research done in the NMR spectroscopy group at the University of Oulu, and the simulations in this study were carried out by using computational resources provided by the CSC – IT Center for Science, namely the Taito supercluster located in Kajaani, Finland. It was a privilege to work on my thesis as a part of consecutive summer jobs in NMR; I would like to thank my supervisors Perttu Lantto and Juha Vaara for the opportunity, not to mention patience as the thesis project dragged on... I would also like to give a big shout-out to Jouni Karjalainen, the programmer of the simulation code I forked for this study, for patiently answering my walls of questions, be it about the simulation code, data analysis, error estimation, anything.

A thesis doesn't exist in a vacuum. For this particular work, previous Master's theses with related subjects have been an invaluable source for inspiration and reference, namely the ones by Jouni, Tapio, Petri and Perttu L. This thesis would be a lot messier without your work, so I would like to thank you for paving the way. Also, special thanks to dear Siiri for meticulously proofreading the thesis and rooting out any Finnish-isms left! Finally, I'd like to give a shout-out to my friends for, well, being there for me: Siiri, Emmu, Pete, Arttu, Tuomas, Tapio, the people at Tuutortupa, Koppi, Muumitalo and everywhere in between — you made my university years all the more memorable.

One more thing; if you haven't studied molecular quantum mechanics (or physics in general, for that matter), and are about to turn the page — well, good luck!

Chapter 1

Introduction

Azobenzene is a broadly studied molecule that forms liquid crystal phases under right circumstances [1]. It has two distinctive geometries, a flat *trans* and a bent *cis* isomer. Through a reversible photoisomerisation process, the molecule can be forced to transform from one isomer to another. Because only the *trans* isomer forms a nematic liquid crystal phase at room temperature, this process additionally triggers a phase transition. Unique qualities of azobenzene and its derivatives are a key to multiple practical applications, such as using polymeric azobenzene films as an image storage medium.

In this study, a bulk of azobenzene molecules is modelled by using a coarse-grained potential and then studied by means of Monte Carlo simulations. The primary goal of the study is to identify liquid crystal phases of *trans*- and *cis*-azobenzene by building phase diagrams from the simulation results. A Fortran 2008 simulation code PTGBCYL by J. Karjalainen was expanded upon and used in the simulations. A coarse-grained model for *cis*- and *trans*-azobenzene was created by fitting a two-site Gay–Berne model [2] into quantum-chemical azobenzene dimer interaction energies. After being implemented into the simulation software, the model was used to simulate a bulk of 2000 azobenzene molecules at constant pressure and temperature across a wide temperature range.

The following four Chapters 2–5 illustrate the theoretical concepts and methods behind this study. In the rest of the thesis, Chapters 6–9, the practical implementation of these methods are explored and the simulation results are discussed. The phase structures of the two isomers are presented along with thermodynamic quantities and structural parameters. The results are then compared into earlier liquid crystal simulation studies with similar single-site Gay–Berne parametrisations and additional two-site Gay–Berne parametrisations. As a result, a nematic liquid crystal phase was identified in the *trans* isomer. The *cis* isomer findings were somewhat unsatisfactory, which might provide suggestions for future simulation studies.

Chapter 2

Molecular electronic structure

Electronic structure of molecules can be solved computationally with various quantum-chemical (QC) methods [3]. The problem is usually formulated as finding the many-electron wavefunction Ψ of the system under inspection by solving its time-independent electronic Schrödinger equation $\hat{H}\Psi = E\Psi$. In an *ab initio* ("from the beginning") calculation, the solution is formed based on fundamental physical principles alone. In practice, Ψ is often approximated with a trial wavefunction which can then be iterated upon with the Hartree-Fock (HF) method. As the complexity of the system grows, *ab initio* methods tend to become computationally expensive — this can be countered with semiempirical methods, where parts of the calculations are furnished with parameters based on experimental data. As a sidenote, there is also an alternative approach to the HF method for solving the electronic structure; density functional theory (DFT), where the electronic structure is represented by electronic density instead of a many-electron wavefunction. In this chapter, I will introduce various wavefunction-based computational methods, emphasising those that are used to build the atomistic models and dimer configurations of *cis*- and *trans*-azobenzene in Chapter 6. Theory regarding this subject is mainly from the 5th edition of *Molecular Quantum Mechanics* by Atkins and Friedman [3].

2.1 Hartree-Fock self-consistent field method

2.1.1 Trial wavefunction

Variational electronic structure-solving methods are, as their name suggests, based on the *variational principle*. These methods begin with a trial wavefunction Ψ chosen to depict a system with a known Hamiltonian $\hat{H}$. Its energy eigenvalue is defined as a

Rayleigh ratio, which is an upper bound for the ground state energy of the system, or

$$E_0 \leq E(\Psi) = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \tag{2.1}$$

where $E_0 = E(\Psi)$ if and only if the chosen trial wavefunction is the *true* wavefunction of the system. This information transforms the problem of finding the correct wavefunction to a problem of minimising the Rayleigh ratio.

The approximate N_e-electron wavefunction Ψ can be constructed from one-electron wavefunctions ψ_m with the *Slater determinant*

$$\Psi(\boldsymbol{x}) = \Psi(1, 2, \ldots, N_e) = \frac{1}{\sqrt{(N_e!)}} \begin{vmatrix} \psi_a(1) & \psi_b(1) & \cdots & \psi_z(1) \\ \psi_a(2) & \psi_b(2) & \cdots & \psi_z(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_a(N_e) & \psi_b(N_e) & \cdots & \psi_z(N_e) \end{vmatrix}. \tag{2.2}$$

Configuration state function (CSF) is a symmetry-adapted linear combination of Slater determinants. Using this determinant as the trial wavefunction ensures that the Pauli exclusion principle[1] is satisfied. The expression is generally simplified to

$$\Psi(\boldsymbol{x}) = (N_e!)^{-1/2} \det |\psi_a(1)\psi_b(2) \ldots \psi_z(N_e)| = ||\psi_a \psi_b \ldots \psi_z||. \tag{2.3}$$

2.1.2 Fock operator

In the original Hartree technique [4], the Coulomb potential experienced by an electron is represented by an averaged potential of a single effective charge, centred in the nucleus of the atom the electrons orbit. (This is known as the *central-field approximation*). Hartree-Fock (HF) self-consistent field method [5] improves this by adding a non-classical electron exchange term to the system Hamiltonian. This term contrastingly cannot be represented as a point charge. These two terms are present in the *Fock operator* $\hat{f}_1$ which is used to solve energy eigenvalues for individual orbital wavefunctions:

$$\hat{f}_1 \psi_m(1) = \varepsilon_m \psi_m(1) \tag{2.4}$$

Here, the eigenvalue ε_m is the *one-electron orbital energy*. The Fock operator consists of the one-electron Hamiltonian h_1, the Coulomb operator $\hat{J}_n$ and the exchange operator $\hat{K}_n$:

$$\hat{f}_1 = \hat{h}_1 + \sum_n [2\hat{J}_n(1) - \hat{K}_n(1)] \tag{2.5}$$

[1] Pauli exclusion principle states that two fermions, e.g., electrons, cannot occupy the same quantum state: they cannot have an identical set of quantum numbers.

These two operators J_n and K_n operate on the wavefunction in the following manner:

$$\hat{J}_n(1)\psi_m(1) = j_0 \int \psi_n^*(2)\frac{1}{r_{12}}\psi_m(1)\psi_n(2)d\tau_2 \tag{2.6}$$

$$\hat{K}_n(1)\psi_m(1) = j_0 \int \psi_n^*(2)\frac{1}{r_{12}}\psi_n(1)\psi_m(2)d\tau_2 \tag{2.7}$$

Here r_{12} is the distance between electrons 1 and 2, and $j_0 = e^2/4\pi\epsilon_0$. The Coulomb operator describes the Coulombic repulsion and the exchange operator the effects of spin correlation, i.e., the repulsion between two same-spin particles.

2.1.3 Roothaan equations

Because the wavefunctions, i.e., the solutions to the Schrödinger Equation, are present in the Fock operator, Eq. (2.4) cannot be solved analytically. Thus, a trial wavefunction is needed as a starting point for a numerical iterative procedure. As the complexity of a molecular system grows, solving the equation quickly becomes too complex a task, so a modification is due. In an approach proposed by C.C.J. Roothaan and G.G. Hall, molecular orbitals were to be represented as a linear combination of atomic orbitals (LCAO), or

$$\psi_m = \sum_{o=1}^{N_o} c_{om}\chi_o. \tag{2.8}$$

The atomic orbitals χ_o form the *basis set* for the calculation. Now the molecular electronic structure is represented by a finite number of basis functions, and the Schrödinger equation can be restructured into a set of N_o simultaneous *Roothaan Equations*

$$\sum_{o=1}^{N_o} F_{o'o}c_{om} = \varepsilon_m \sum_{o=1}^{N_o} S_{o'o}c_{om}, \tag{2.9}$$

where

$$F_{o'o} = \int \chi_{o'}^*(1) f_1 \chi_o(1) d\tau_1 \tag{2.10}$$

and

$$S_{o'o} = \int \chi_{o'}^*(1)\chi_o(1)d\tau_1. \tag{2.11}$$

As the orbitals themselves are present in (2.10) and (2.11), these equations have to be solved iteratively. In an iteration cycle, an initial set of wavefunctions ψ_m (that is, atomic orbitals χ_o and coefficients c_{om}) are used to construct operators (2.10) and (2.11) and then solve Eqs. (2.9). Thus, new wavefunctions are acquired. This iteration is repeated until new wavefunctions do not significantly differ from the ones in the beginning of the cycle, and the solution is said to be *self-consistent*.

2.2 Electron correlation

When using the HF method to calculate the wavefunction for a given electron, the effect of all the other particles in the system is averaged into a single field. This averaging disregards quantum mechanical effects on the electron distribution caused by instantaneous Coulombic electron-electron interactions, often referred to as *electron correlation*. Correlation energy is defined as the difference between the actual energy and the HF energy of the system, $E_{corr} = E - E_{HF}$. Correlation energy itself can be divided into dynamic (short-ranged) and nondynamical (long-ranged) effects [6]. In this section, various approaches to estimate E_{corr} are introduced.

2.2.1 Configuration interaction

Let us consider a ground-state wavefunction constructed out of spin orbitals ψ_i with the Slater determinant notation defined in section 2.1.1:

$$\Psi = ||\psi_a \psi_b \dots \psi_l \psi_m \psi_n \dots \psi_z||. \tag{2.12}$$

If an electron in spin orbital m is excited to orbital p, the wavefunction becomes

$$\Psi_m^p = ||\psi_a \psi_b \dots \psi_l \psi_p \psi_n \dots \psi_z||. \tag{2.13}$$

Similarly, if electrons in orbitals m and n are excited to p and q, respectively, the wavefunction is denoted as

$$\Psi_{mn}^{pq} = ||\psi_a \psi_b \dots \psi_l \psi_p \psi_q \psi_o \dots \psi_z||. \tag{2.14}$$

The exact ground-state wavefunction is formulated as a linear combination of all possible Slater determinants, or

$$\Psi = c_0 \Psi_0 + \sum_{a,p} c_a^p \Psi_a^p + \sum_{\substack{a<b \\ p<q}} c_{ab}^{pq} \Psi_{ab}^{pq} + \sum_{\substack{a<b<c \\ p<q<r}} c_{abc}^{pqr} \Psi_{abc}^{pqr} + \dots . \tag{2.15}$$

This variational approach is called *configuration interaction* (CI). If all possible excitations are considered, we acquire the exact non-relativistic ground-state energy within the Born-Oppenheimer approximation. This is called *full CI* and is straightforward for simple molecules, but becomes extremely demanding when complexity of the system, and thus, the number of all possible excitations increases. One way to reduce computational load is to only include singly and doubly excited molecules (SDCI). Another way is to abandon variational methods altogether; this is explored in the following sections.

2.2.2 Møller-Plesset perturbation theory

Electron correlation can be approximated with methods based on perturbation theory. Rayleigh-Schrödinger perturbation theory starts with an unperturbed Hamiltonian operator and introduces a small perturbation to it. Methods involving perturbation theory (PT) are non-variational, so the acquired energy will not be an upper bound for the energy any more. Unlike truncated CI, however, PT methods are *size-consistent*, which means the accuracy of the calculation is independent of system size.

Møller-Plesset perturbation theory (MPPT) [7] considers the HF configuration as a reference state and appends electron correlation as a Rayleigh-Schrödinger-type perturbation. The unperturbed zero-order Hamiltonian $\hat{H}^{(0)} = \hat{H}_{\mathrm{HF}}$ is constructed as a sum of one-electron Fock operators defined in Eq. (2.5). Its eigenfunction is the ground-state wavefunction Ψ_0. The corresponding eigenvalue, zero-order energy, is denoted by $E_0^{(0)}$. When a first-order perturbation is applied, the Hamiltonian of the system becomes

$$\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)} = \hat{H}_{\mathrm{HF}} + \hat{H}^{(1)}. \tag{2.16}$$

The first-order perturbation for electron i is

$$\hat{H}^{(1)}(i) = \hat{H}(i) - \hat{H}_{\mathrm{HF}}(i) = j_0 \sum_{j \neq i} \frac{1}{r_{ij}} - \sum_n [2J_n(i) - K_n(i)]. \tag{2.17}$$

The corresponding eigenvalue is the first-order energy $E_0^{(1)}$, and the eigenvalue of $\hat{H}$ is the sum of zero-order and first-order energies. Moreover, the second-order correction to the energy is

$$E^{(2)} = \sum_{J \neq 0} \frac{\langle \Psi_J | \hat{H}^{(1)} | \Psi_0 \rangle \langle \Psi_0 | \hat{H}^{(1)} | \Psi_J \rangle}{E_0^{(0)} - E_J^{(0)}}. \tag{2.18}$$

Calculations involving higher-order perturbations are increasingly complex, and the most used MPPT method is the second-order correction-denoted *MP2*.

2.2.3 Coupled-cluster method

Another method of estimating electron correlation is the theory of *coupled clusters* (CC) [8] which is based on operating the HF reference state Ψ_0 with an exponential cluster operator $e^{\hat{C}}$. The exponential operator is constructed as a series expansion

$$e^{\hat{C}} = 1 + \hat{C} + \frac{1}{2!}\hat{C}^2 + \frac{1}{3!}\hat{C}^3 + \ldots, \tag{2.19}$$

where the cluster operator $\hat{C}$ is a sum of one-electron, two-electron, ..., and N-electron excitation operators:

$$\hat{C} = \hat{C}_1 + \hat{C}_2 + \cdots + \hat{C}_N. \tag{2.20}$$

The i-electron excitation operators are applied to the HF wavefunction like this:

$$\hat{C}_1\Psi_0 = \sum_{a,p} t_a^p \Psi_a^p,\ \hat{C}_2\Psi_0 = \sum_{a,b,p,q} t_{ab}^{pq}\Psi_{ab}^{pq}, \ldots \tag{2.21}$$

Here, t_a^p are the single-excitation amplitudes, t_{ab}^{pq}, the double-excitation amplitudes, etc. These representations can be now inserted into the cluster operator (2.19) and the Schrödinger equation for the trial CC wavefunction $\Psi = e^{\hat{C}}\Psi_0$ is solved iteratively [9].

Like MPPT, the CC theory is size-consistent. In widely used approximations, only some of the first terms in (2.20) are taken into account. In CCSD (coupled-cluster singles and doubles), $\hat{C} \approx \hat{C}_1 + \hat{C}_2$, in CCSDT (singles, doubles and triples) $\hat{C} \approx \hat{C}_1 + \hat{C}_2 + \hat{C}_3$ and so on [10]. An approximate form of CCSD is called CC2 [11].

2.2.4 Spin scaling

To improve accuracy, correlation energy from CC or MP calculations can be modified with semiempirical scaling factors. Same-spin (Spin component scaling, SCS) [12] or opposite-spin (Scaled opposite-spin, SOS) [13] scaling factors can be added to the spin part of the correlation energy [14]. Correlation energy is divided into parallel (same-spin) and antiparallel (opposite-spin) components that are scaled separately. The introduction of scaling factors substantially improves the accuracy of MP2 calculation and extends its usability as a quantum-chemical method.

2.3 Basis sets

Molecular orbital basis sets consist of element-specific basis functions. To acquire the exact HF energy, one would have to utilise infinite number of basis functions — this is called the *HF limit*. Real-world calculations are limited to finite basis sets, and the related error is called the *basis-set truncation error*. To optimise computational resources, basis functions must be cleverly chosen to acquire a result sufficiently close to the HF limit, simultaneously utilising as few basis functions as possible.

2.3.1 Slater- and Gaussian-type orbitals

The most common basis functions are Slater-type orbitals (STO) [15] and Gaussian-type orbitals (GTO) [16]. The radial part of a STO is of the form

$$R_{\mathrm{STO}}(r) = N_{\mathrm{STO}}(l,\alpha)\ r^l\ e^{-\alpha r}. \tag{2.22}$$

Here N is the normalising factor related to the function, l the angular momentum quantum number and α a positive exponent. In similar terms, the radial part of a

GTO can be written as

$$R_{\mathrm{GTO}}(r) = N_{\mathrm{GTO}}(l, \alpha)\ r^l\ e^{-\alpha r^2}. \tag{2.23}$$

The origin of these functions is positioned in the atomic nucleus, and r is the distance measured from there. A distinguishing feature of GTOs is an ability to split its radial part into Cartesian components:

$$e^{-\alpha r^2} = e^{-\alpha |x|^2} e^{-\alpha |y|^2} e^{-\alpha |z|^2}. \tag{2.24}$$

This makes calculating integrals more straightforward, as three-dimensional integrals can be split into three one-dimensional integrals [17]. The Cartesian form of the Gaussian is

$$g_{ijk}(\boldsymbol{r}) = N_{\mathrm{GTO}}(l, \alpha)\ x^i y^j z^k\ e^{-\alpha r^2}, \tag{2.25}$$

where orbitals s, p, d, etc. correspond to the sum $i+j+k$ being 0, 1, 2, etc, respectively. GTOs have another important advantage over STOs, too: the product of two Gaussians is a new Gaussian whose origin is the weighted average of the original origins [17], positioned on a line connecting the two original Gaussians. This property can be used to reduce the number of functions in multicenter integral calculations, and thus reduce the computational complexity altogether. It should be noted, however, that STOs are better than GTOs at depicting the electronic wavefunction near the atomic nucleus. Because of that, more GTOs are needed to acquire accuracy similar to STOs. Another way to improve accuracy is to use *contracted Gaussian functions* χ_0, which are linear combinations of primitive Gaussians g_i, or

$$\chi_0 = \sum_i d_{0i} g_i, \tag{2.26}$$

where d_{0i} is a contraction coefficient. The final molecular orbitals are constructed as linear combinations of these contracted Gaussians.

2.3.2 Basis set types

The simplest form of a molecular basis is called the *minimum basis*, where one function represents exactly one atomic orbital. More accurate results are achieved by using more functions per one orbital. Two functions make for a Double-zeta (DZ) basis set, three functions a Triple-zeta basis, four functions a Quadruple-zeta basis, and so on (TZ, QZ, 5Z, ...). Split-valence (SV) type basis sets are a compromise between the minimum basis and multiple-zeta bases, where every inner shell consists of one basis function, but valence orbitals are depicted by 2, 3, 4, ... functions.

To consider atomic orbital polarisation present in molecular bonds, an additional

orbital is included in the basis set — one with a higher quantum number l than in the orbitals that are occupied in the system. This is denoted by letter P at the end of the basis set name: e.g., TZV basis of carbon includes s and p orbitals; TZVP includes s, p, and d; TZVPP includes s, p, d, f, etc.

Basis sets can be *correlation consistent* (cc), which means the basis sets converge systematically towards the complete basis set limit, which is the extrapolated energy limit that would be acquired with an infinitely large basis set. *Augmented* basis sets (aug) include auxiliary diffuse functions for added accuracy. Using an auxiliary basis set expansion can significantly reduce computational cost in the case of multicentre integrals [18]. This technique is referred to as "resolution of the identity" (RI), and is often used to accompany electron correlation calculations such as MP2 ("RI-MP2") [19].

The most notable basis set types are Ahlrichs [16], Dunning [20], and Pople [21]. The basis set notation explained above applies specifically to Ahlrichs basis sets, which are used in this study. Dunning basis sets use a similar notation, while the older Pople basis sets are named in a shorter fashion — some exemplar basis sets are listed in Table 2.1 [14].

Table 2.1: Some basis functions listed. Basis sets beginning with "def2" are from the basis set library of the TURBOMOLE software. Notably, the def2-TZVPP basis is used in the quantum-chemical calculations of this study.

Basis set	Type	Explanation
def2-SV(P)	Ahlrichs	**S**plit-**v**alence
def2-TZVP	Ahlrichs	**T**riple-**z**eta split-**v**alence plus **p**olarisation
def2-TZVPP	Ahlrichs	same as above, but doubly polarised
cc-pVQZ	Dunning	**c**orrelation-**c**onsistent **p**olarised **v**alence **q**uadruple-**z**eta
aug-cc-pVQZ	Dunning	same as above, but augmented
6-31G*	Pople	1 contracted core orbital from **6** Gaussians
		2 contracted valence orbitals from **3** and **1** Gaussians
		plus polarisation (*****)

2.3.3 Basis-set superposition error

Interaction energy in a dimer system of molecules A and B is defined as [22]

$$\Delta E_{\mathrm{AB}} = E_{\mathrm{AB}}(\mathrm{AB}) - E_{\mathrm{A}}(\mathrm{A}) - E_{\mathrm{B}}(\mathrm{B}), \tag{2.27}$$

where the subscript denotes which molecules are included in the basis; the inspected system is in parentheses. Basis-set superposition error (BSSE) manifests itself when the molecules are close enough to one another so that their electron clouds start to overlap. When it happens, the unoccupied basis functions of molecule A lower the energy of molecule B and vice versa. The shift in energy is not a problem per se; the problem is that the shift *only* happens when molecules A and B are close — this distorts the overall potential surface. BSSE can be corrected with *counterpoise correction* [23]: monomer energies are calculated with a full dimer basis, i.e., $E_{\mathrm{AB}}(\mathrm{A})$, and the resulting counterpoise-corrected energy equals

$$\Delta E_{\mathrm{AB}}^{\mathrm{CP}} = E_{\mathrm{AB}}(\mathrm{AB}) - E_{\mathrm{AB}}(\mathrm{A}) - E_{\mathrm{AB}}(\mathrm{B}). \tag{2.28}$$

The amount of BSSE is now

$$E_{\mathrm{AB}}^{\mathrm{BSSE}} = \Delta E_{\mathrm{AB}}^{\mathrm{CP}} - \Delta E_{\mathrm{AB}}. \tag{2.29}$$

Chapter 3

Liquid crystals

3.1 Liquid crystal terminology

A solid crystal consists of particles positioned in a periodically repeating structure with long-range order [24]. The order present in a crystal lattice can be translational order (particles occupying specific positions), or orientational order (particles pointing to specific directions). Contrastingly, in a conventional liquid, there is no long-range order at all: the particles are randomly oriented and positioned. Liquid crystals (LC) occupy the phase space somewhere between liquids and crystals. LC particles, or *mesogens*, can have crystal-like order, but to a much lesser extent (and with a shorter range), and some of the 'flowiness' of a liquid phase is apparent: the particles are not strictly confined to their initial positions and can experience diffusion.

The most common LC-forming molecules are rodlike, or *calamitic*. Another common type is a flat, disc-like molecule with a rigid centre; these are called *discotic*. Many liquid crystals are *thermotropic*, which means that the LC phase only appears in a certain temperature range (given a suitable pressure). Some molecules only form LC phases when mixed with a right solvent; these compounds are called *lyotropic* and include phospholipids and soap. The main focus in this study is on calamitic, thermotropic liquid crystals.

3.2 Liquid crystal phases

Liquid crystal phases can be characterised by the nature of their ordering. The phase present in conventional liquids and gases, the *isotropic phase* (I), is devoid of any order, whereas liquid crystal phases are anisotropic. When a crystalline solid (Cr) is heated, thermal kinetic energy of the particles increases until the regular structure starts to break down. The crystal phase transitions into isotropic liquid, and for some compounds, this can happen via an intermediate liquid crystalline mesophase.

The simplest calamitic liquid crystal phase is the *nematic phase* (N) characterised by orientational order, or a preferred orientation of the particles. This is often described with a director $\hat{\boldsymbol{n}}$, a unit vector aligned to the unique principal axis of the orientation tensor $\boldsymbol{Q}$. The tensor is calculated as an average over the liquid crystal particles in the system, and the director can be used to visualise the direction of the orientational average of the particles. If the particle orientations are random in all the other directions than the director, the phase is *uniaxial*. If there are two preferred directions, the phase is *biaxial*. This can manifest itself especially if the particles itself are banana- or V-shaped bent-core particles, but the bent shape is not a necessary condition for the formation of biaxial phases: they can also form in a mixture of calamitic and discotic particles [25]. Nematic phase-forming particles work as a basis for liquid crystal display technology, which utilises materials whose orientation can be adjusted with an electric field [24].

The *smectic phase* (Sm) is more ordered than the nematic phase. There is, in addition to orientational order, some translational order present: the particles are arranged in layers. Smectic phase has many variants. In smectic-A (SmA), the director points to the direction perpendicular to the layer planes. In smectic-C (SmC), director is slightly tilted to some other direction. Other variants of the smectic phase include phases where layered particles have hexagonal positional order: in smectic-B (SmB), the director is not tilted, and in smectic-I (SmI) and smectic-F (SmF) it is. Some phases of interest are pictured in Fig. 3.1.

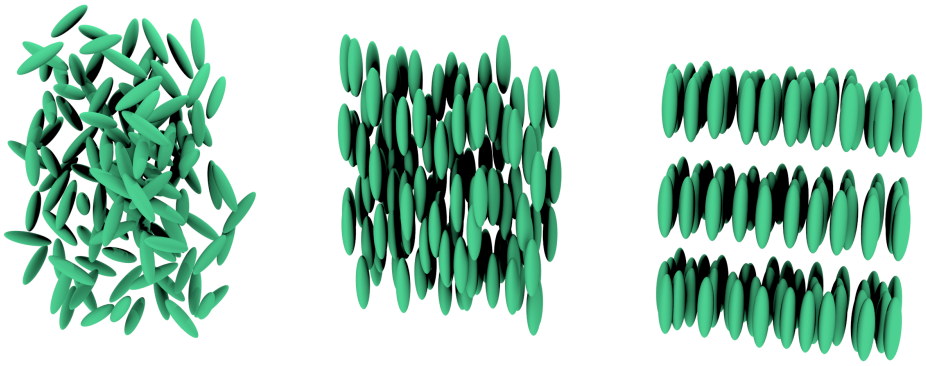


Figure 3.1: From left to right: isotropic, nematic and smectic-A phase of a calamitic mesogen (with exaggerated amount of order for clarity).

LC phases can contain *chirality* (orientation of the director depending on spatial location), especially when the molecule that forms the phase is chiral itself. Calamitic LCs have a chiral variant of the nematic phase where the director rotates around the axis perpendicular to molecular orientation. Tilted smectic phases have chiral variants, too, where the director angle slightly changes when moved from one layer to the next.

3.3 Liquid crystal phase determination

3.3.1 Orientational order parameter

Order of molecular orientation can be examined with *orientational order parameters*. In this study, the second-rank orientation tensor $\boldsymbol{Q}$ [26–28] is used, and its components for an N-particle system are calculated with

$$Q_{\alpha\beta} = \frac{1}{2N}\sum_{i=1}^{N} 3\left(\hat{u}_{i\alpha}\hat{u}_{i\beta} - \delta_{\alpha\beta}\right);\ \alpha,\beta = x,y,z. \tag{3.1}$$

Here, the components $\hat{u}_{i\alpha}$ ($\alpha = 1,2,3$) comprise the unit vector $\hat{\boldsymbol{u}}_i$ that is aligned to the longitudinal axis of particle i, and $\delta_{\alpha\beta}$ is the Kronecker delta. Eigenvalues and eigenvectors of this tensor are of particular interest: the orientational order parameter P_2 is the largest eigenvalue of $\boldsymbol{Q}$ and the liquid crystal director $\hat{\boldsymbol{n}}$ is the eigenvector that corresponds to it[1]. The values of P_2 range from zero to one, zero describing the complete lack of orientational order, and one a system whose all particles align along a shared axis. In simulations, P_2 has a nonzero (albeit small) value even in isotropic phases due to a finite system size [29]. For nematic phases, $P_2 \approx 0.3 - 0.7$ [30].

If the director is aligned along the z-axis, the orientational order parameter becomes

$$P_2 = Q_{33} = \frac{1}{2}\left\langle 3\hat{u}_{iz}^2 - 1\right\rangle, \tag{3.2}$$

where the brackets $\langle \dots \rangle$ denote an average over all the particles. The final, presented orientation parameter is additionally averaged over multiple simulation snapshots.

The choice of the vector $\hat{\boldsymbol{u}}_i$ in the $\boldsymbol{Q}$-tensor is trivial for uniaxial particles. For biaxial bent-core particles, $\hat{\boldsymbol{u}}_i$ is calculated as a difference of the two unit vectors that determine the particle orientation, or by using notation in Fig. 4.4, $\hat{\boldsymbol{u}}_i = \hat{\boldsymbol{u}} - \hat{\boldsymbol{v}}$. Additionally, in the biaxial nematic phase, the $\boldsymbol{Q}$-tensor can be written as

$$\boldsymbol{Q} = \begin{pmatrix} Q_{11} & 0 & 0 \\ 0 & Q_{22} & 0 \\ 0 & 0 & Q_{33} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}S + T & 0 & 0 \\ 0 & -\frac{1}{2}S - T & 0 \\ 0 & 0 & S \end{pmatrix} \tag{3.3}$$

Here $S = P_2$ and T is the *measure of biaxiality*. From here, T can be solved with $Q_{33} = S$ and either Q_{11} or Q_{22}:

$$T = Q_{11} + \frac{1}{2}Q_{33} = -Q_{22} - \frac{1}{2}Q_{33} \tag{3.4}$$

[1]To be exact, P_2 is the Legendre polynomial $P_2(x) = \frac{1}{2}(3x^2 - 1)$, and the orientational order parameter is $\langle P_2(\cos\theta)\rangle$, where θ is the director angle [28]. When θ is measured with respect to z-axis, the parameter becomes $\langle P_2(\hat{u}_{iz})\rangle$. See Eq. (3.2) for notation used in this study.

3.3.2 Translational order parameter

Translational order parameter τ_1 denotes how layered the particles are in the smectic phase and is defined for an N-particle system by [29]

$$\tau_1 = \max\left|\frac{1}{N}\sum_{i=1}^{N}\exp(2\pi i r_{i\parallel}/d)\right|, \tag{3.5}$$

where $r_{i\parallel}$ is the dot product $\boldsymbol{r}_i \cdot \hat{\boldsymbol{n}}$. Here $\boldsymbol{r}_i$ is the position vector pointing to the centre of particle i and $\hat{\boldsymbol{n}}$ is the already familiar liquid crystal director. To find the maximum of the sum, the distance between layers d is varied in a predetermined interval. It is noteworthy that the exponentiation returns a complex number; thus the closer the complex numbers are in phase with each other, the larger the sum [30]. The absolute value in the equation denotes the modulus of this complex number. For isotropic and nematic phases, the layer structure is nonexistent and $\tau_1 \approx 0$. The maximum value $\tau_1 = 1$ denotes a perfectly regular layer structure omnipresent in the system.

3.3.3 Bond orientational order parameter

The hexagonal order parameter a.k.a. *bond orientational order parameter* measures how hexagonally ordered the particles in a smectic layer are. For a single particle i, local hexagonal order is given by [29]

$$\psi_6(\boldsymbol{r}_i) = \frac{\sum_j w(r_{ij})\exp\left(6i\theta_{ij}\right)}{\sum_j w(r_{ij})} \tag{3.6}$$

as a summation over other molecules $j \neq i$. Here θ_{ij} is the angle of the "bond" between particles $i,\ j$ in a layer. In the perfectly hexagonal case it equals 60°. More precisely, it is the angle between the $i-j$ position vector $(\boldsymbol{r}_i - \boldsymbol{r}_j)$ projected onto the smectic layer and a fixed reference vector on the smectic layer. The function $w(r_{ij})$ returns one or zero for a particle pair i, j. By cleverly choosing $w(r_{ij})$ one can decide which molecules are considered to be "close enough" to each other to be taken into account in the summation. The bond orientation parameter of the whole system is the average of the *local* bond orientation order parameter:

$$\psi_6 = |\langle\psi_6(\boldsymbol{r}_i)\rangle| \tag{3.7}$$

3.3.4 Radial distribution function

Radial distribution function $g(r)$ (RDF, also known as *pair distribution function* and *pair correlation function*) gives information about the distribution of interparticular distances, and thus describes the structure of the system. It tells the probability of finding a particle j (a "neighbour") at a distance r_{ij} from a reference particle i, compared to ideal gas distribution of the same density [31]. In mathematical terms, the spherically symmetric pair distribution function is

$$g(r) = \frac{1}{NN_{id}} \sum_i \sum_{j \neq i} \delta(r, r_{ij}), \tag{3.8}$$

where

$$\delta(r, r_{ij}) = \begin{cases} 1, \text{ if } r - \delta r/2 < r_{ij} \leq r + \delta r/2 \\ 0, \text{ otherwise} \end{cases} \tag{3.9}$$

The delta function makes sure that only the particles inside the spherical shell are considered: Essentially, the summation tells how many particles j are located in a particle i-centered spherical shell of radius r and thickness δr. This is normalised with the total number of atoms N and the average number of atoms in a spherical shell of ideal gas with the same density and shell radius, or

$$N_{id} = \frac{4\pi\rho}{3} \left[\left(r + \frac{\delta r}{2} \right)^3 - \left(r - \frac{\delta r}{2} \right)^3 \right]. \tag{3.10}$$

Here, density ρ is defined as a number density $\rho = \frac{N}{V}$ (number of particles N per volume V).

A schematic RDF representation for gas, liquid and crystal phases for a simple atomistic particle is shown in Fig. 3.2. For every phase, the first peak occurs around the particle diameter σ, or the *first coordination shell*. Peaks at σ, 2σ, 3σ, ... characterise the liquid nature of the phase, and sharp peaks with clear separation distances indicate a crystal lattice structure. Moreover, peaks are more frequent in a crystal than in a liquid. This indicates longer-range order, which is diminished in liquid through diffusion. A gaseous state is signified by a single peak after which the graph slowly decays into unity [32].

When calculating $g(r)$ in practice from simulation results, neighbours of the particles are sorted into bins, or histograms, with a chosen distance precision δr. Additionally, it is noteworthy to mention that when dealing with a simulation with periodic boundary conditions (see section 5.6), the distance between two particles r_{ij} is not unique. Thus, it is important to only consider the shortest distance. This is known as the *minimum image convention* [31].

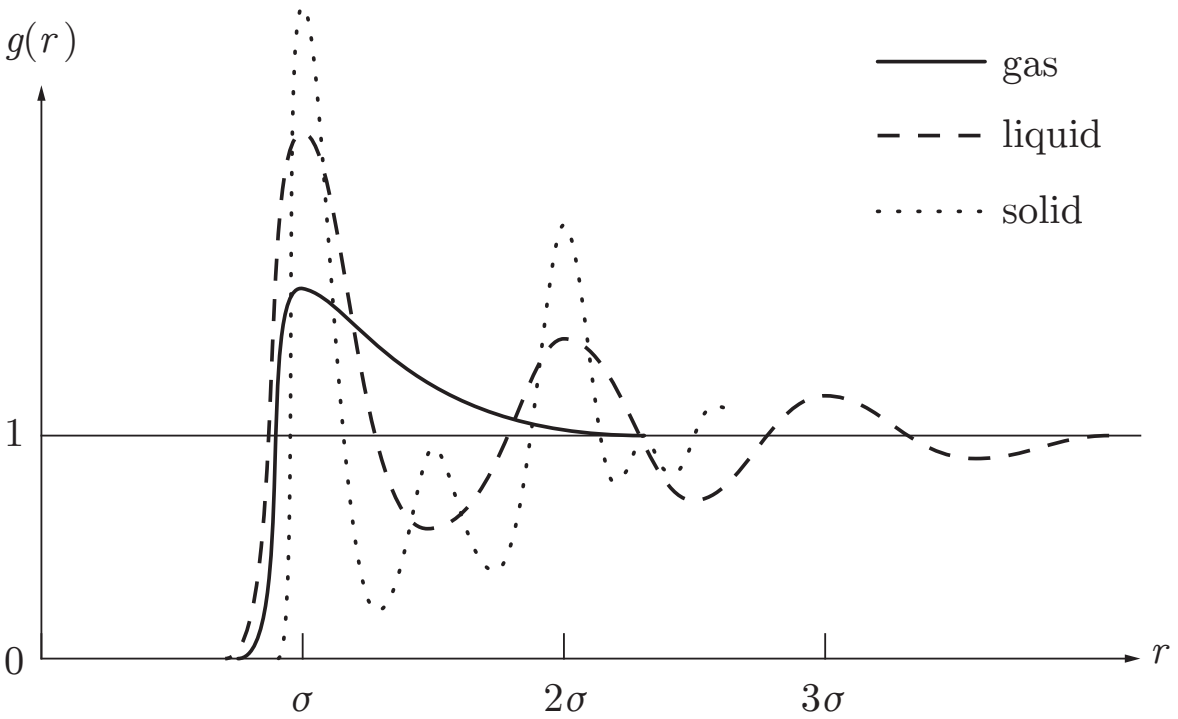


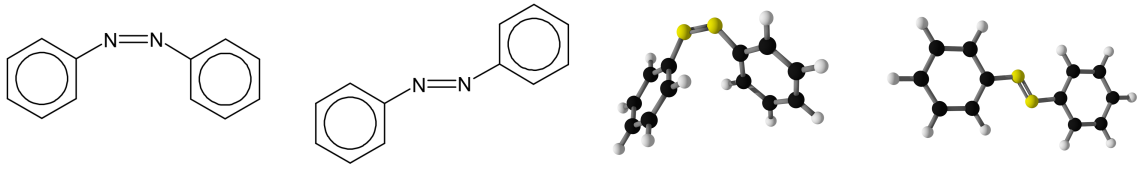
Figure 3.2: Schematic radial density functions of three phases: gas, liquid and solid. Here, σ is the particle Van der Waals diameter, defined roughly as the closest distance two atoms can get to each other [32].

Chapter 4

Coarse-grained model for azobenzene

4.1 Azobenzene molecule

The azobenzene molecule consists of two benzene rings conjoined by a nitrogen-nitrogen double bond, or an azo group. The two isomers of azobenzene, *cis*-azobenzene (CAB) and *trans*-azobenzene (TAB), have distinct geometries. The *trans* isomer is planar, but the benzene rings of the *cis* form are tilted and slightly twisted, as seen in Fig. 4.1.



(a) **(b)**

Figure 4.1: (a) Structural formulae and (b) quantum-chemically optimised geometries for *cis*- (left) and *trans*-azobenzene molecules.

Azobenzene and its derivatives are a broadly studied group of molecules [33, 34]. One especially interesting property is rapid, reversible photoisomerisation: ultraviolet light can be used to induce a *trans*-to-*cis* transition undergoing in 10 μs [1]. Visible blue light induces a *cis*-to-*trans* transition. Additionally, *cis*-azobenzene relaxes thermally to the energetically lower *trans* state. This back-isomerisation happens slowly at room temperature but it is especially effective at temperatures past 100 °C [1]. A schematic energy graph for the photoisomerisation is seen in Fig. 4.2. It was earlier suggested that the mechanisms for the state transition include two pathways — an in-plane inversion and a rotation. Later studies have shown that both isomerisation processes are purely rotational [33].

The *trans* form of several azobenzene derivatives is known to form a nematic LC

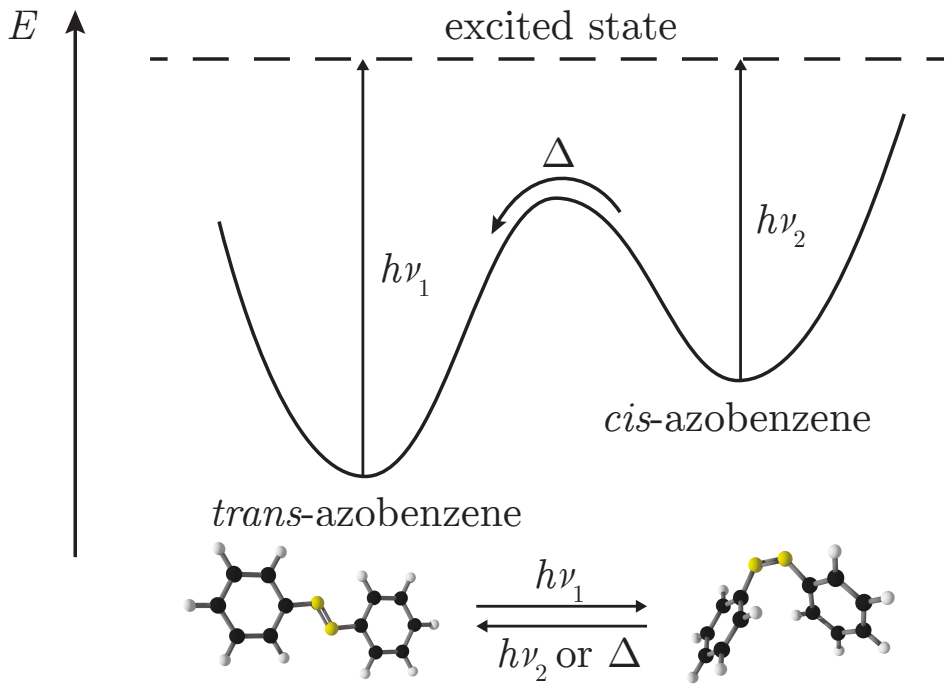


Figure 4.2: A schematic energy graph for the state transitions between *cis*- and *trans*-azobenzene. Photon with an energy $h\nu_1$ or $h\nu_2$ induces a transition via an excited state, and additionally *cis* form relaxes thermally to *trans* form (Δ).

phase, but no LC phases have been found for the *cis* isomer [1, 35]. This means the photoisomerisation effect can be used to trigger a nematic-to-isotropic phase transition, and vice versa. At higher temperatures, the nematic phase undergoes a transition into isotropic phase, and at lower temperatures, it solidifies into a highly ordered crystal [36]. Some azobenzene derivates can even form a smectic-A phase and show temperature-dependent molecular biaxiality [37].

To say azobenzene has many applications is an understatement. Firstly, azobenzene is a derivative of diazenes, which significantly absorb light, so it can be used as a dye. Photoisomerisation opens up possibilities for a myriad of important applications. Polymeric azobenzene LC films can be used as an image storage material: starting in the nematic phase, if a *trans–cis–trans* transition cycle is performed, the resulting *trans* molecules remain in the isotropic phase, even if temperature is kept under a certain threshold T_g. Reorientation to nematic phase happens when the temperature is raised to above T_g. At lower temperatures these LC films can be used for image storaging by means of laser irradiation combined with photomasking [1]. Additionally, azobenzene-containing polymers (azo-polymers) have been used to build alignment layers for fluorescent polymers [38], high-density data storage [39], reversibly switchable superhydrophobic surfaces [40], etc. Other applications of photoisomerisation include light-driven motion of macroscopic liquid droplets on a flat solid surface [41] and artificial muscles based on azobenzene elastomers [42].

4.2 Coarse-grained model

4.2.1 Coarse-graining

Simulating azobenzene in atomistic detail would require infeasible simulation times due to the number of interactions to be calculated. Thus, the azobenzene phase structure is simulated with a *coarse-grained* (i.e., simplified) model with only two interaction sites per molecule. Coarse-graining is an umbrella term for methods that reduce the complexity of the observed system by substituting a detailed (usually atomistic) model with an approximate one with fewer interaction sites [43, 44]. This leads to new possibilities of simulating systems with longer timescales and larger sizes at the expense of molecular detail. One particular application for coarse-graining is modelling large biomolecules, e.g., proteins [45].

4.2.2 Gay–Berne potential

The Gay–Berne (GB) potential is essentially an anisotropic extension of the spherically symmetric Lennard-Jones (LJ) potential [46], which has an attractive term with exponent 6 and a repulsive term with exponent 12, and is defined as

$$U_{\text{Lennard-Jones}} = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]. \tag{4.1}$$

Here, ϵ is the potential strength, σ the potential contact distance and r_{ij} the distance between the two interacting particles i and j. The Gay–Berne potential was developed in 1980 by Gay and Berne [2], who based it on an earlier overlap potential by Berne and Pechukas [47]. Like the LJ potential, the GB potential is a pair potential that describes attractive and repulsive forces between two interaction sites, i.e., particles. The GB particles have a shape of a prolate or oblate spheroid and are thus axially symmetric. The strength and range of the interaction depend on the orientations of the two interacting particles. Potential energy graphs for some interesting configurations are pictured in Fig. 4.3: side-by-side (SBS), end-to-end (ETE), cross (X) and T. Like the LJ potential, the attractive and repulsive parts have a similar 6-12 form:

$$U_{\text{GB}}(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}) = 4\epsilon(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}) \left[R^{-12} - R^{-6} \right]. \tag{4.2}$$

Instead of the original notation of Gay and Berne, a notation used in a 1999 study by Luckhurst and Bates [29] is applied in this section. This notation utilises a helper

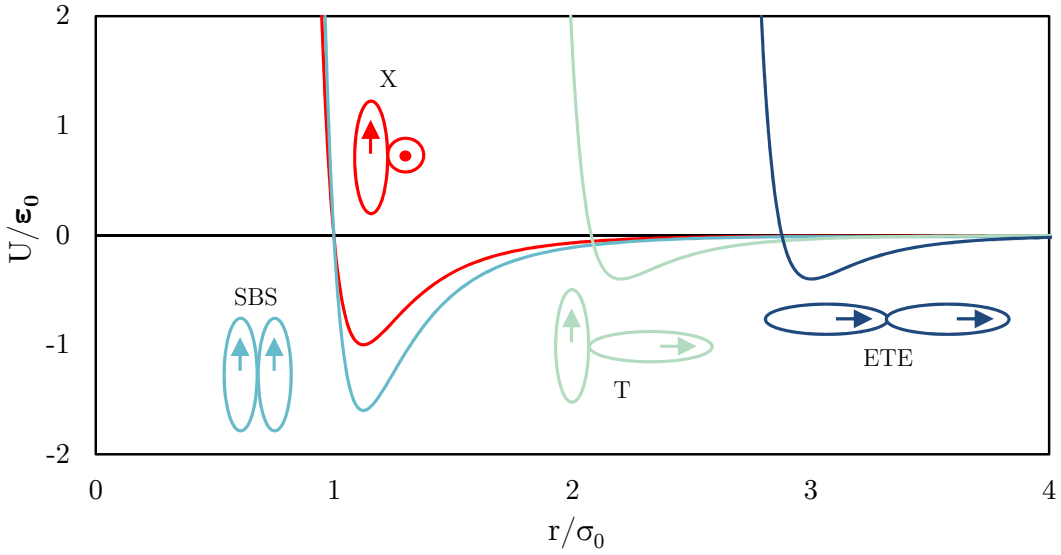


Figure 4.3: Potential energy graphs of the anisotropic Gay–Berne potential in units of ϵ_0 and σ_0 for four two-particle configurations: side-by-side (SBS), end-to-end (ETE), X and T.

function R, or the shifted and scaled separation between uniaxial particles:

$$R = \frac{r_{ij} - \sigma(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}) + \sigma_0}{\sigma_0} \tag{4.3}$$

Here, unit vectors $\hat{\boldsymbol{u}}_i$ and $\hat{\boldsymbol{u}}_j$ represent the directions of the symmetry axes of particles i and j. $\boldsymbol{r}_{ij}$ is the displacement vector between the particles, r_{ij} its length and $\hat{\boldsymbol{r}}_{ij}$ the associated normalised unit vector. In Eqs. (4.2) and (4.3), functions $\sigma(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij})$ and $\epsilon(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij})$ are the *range* and *strength parameters*. The range parameter measures the contact distance between two GB particles in accordance to their orientations and positions:

$$\sigma(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}) = \sigma_0 \left\{ 1 - \chi \left[\frac{(\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij})^2 + (\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij})^2 - 2\chi(\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij})(\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij}2)(\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_j)}{1 - \chi^2(\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_j)^2} \right] \right\}^{-1/2} \tag{4.4}$$

Here, σ_0 is the parameter that defines the unit of contact distance. More specifically, it is the distance between the interaction site centres when the potential $U_{GB} = 0$ in SBS and X configurations. When visualising the molecule, σ_0 is commonly used as the width of the GB particle, or the *equatorial diameter*. In Eq. (4.4),

$$\chi = \frac{\kappa^2 - 1}{\kappa^2 + 1}, \tag{4.5}$$

where the parameter $\kappa = \sigma_{\text{ete}}/\sigma_{\text{sbs}}$ is the ratio of contact distance in ETE and SBS configurations, and can be thought as the particle length-to-breadth ratio. Thus, the length, or *polar diameter* of the particle, can be represented as $\kappa \cdot \sigma_0$.

The strength parameter is defined by a product of two functions, first of which is, rather confusingly, also called ϵ:

$$\epsilon(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}) = \epsilon_0 \epsilon^{\nu}(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j) \epsilon'^{\mu}(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}) \tag{4.6}$$

Here, ϵ_0 is the parameter that scales the potential strength and defines the unit of energy. It is also the potential well depth in the X configuration. The $\epsilon(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j)$ function is the strength parameter for two same-shaped spheroids in the overlap potential by Berne and Pechucas:

$$\epsilon(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j) = \epsilon_0 \left[1 - \chi^2 (\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_j)^2\right]^{-1/2}. \tag{4.7}$$

The ϵ' has a form similar to $1/\sigma^2$ (see Eq. (4.4)):

$$\epsilon'(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{ij}) = 1 - \chi' \left[\frac{(\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij})^2 + (\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij})^2 - 2\chi'(\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{r}}_{ij})(\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{r}}_{ij} 2)(\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_j)}{1 - \chi'^2 (\hat{\boldsymbol{u}}_i \cdot \hat{\boldsymbol{u}}_j)^2} \right], \tag{4.8}$$

where

$$\chi' = \frac{\kappa'^{1/\mu} - 1}{\kappa'^{1/\mu} + 1}. \tag{4.9}$$

An important parameter, $\kappa' = \epsilon_{\text{sbs}}/\epsilon_{\text{ete}}$ is introduced here — it is the ratio of potential strenght in SBS and ETE configurations. Additionally, parameters μ and ν present in Eqs. (4.6) and (4.9) scale the relative strength of (4.7) and (4.8).

To summarise this section, we now have four parameters that are needed to fully control the anisotropy of the GB potential: (κ, κ', μ, ν). Additionally, the units of distance and energy are scaled by parameters (σ_0, ϵ_0). If just identical particles are considered, we can set $\sigma_0 = \epsilon_0 = 1$.

4.2.3 Two-site Gay–Berne potential

The GB potential has been extensively used to study molecular model behaviour [48–50], producing nematic, smectic-A and smectic-B liquid crystal phases. However, the single-site GB potential is not flexible enough to represent the bent-core *cis*-azobenzene. For this reason, a two-site continuation of the Gay–Berne potential (2GB) is introduced here, consisting of two identical GB interaction sites joined from their ends — a single site representing one half of the molecule. By adjusting the bending angle between the two sites, this potential can be used for both *cis*- and *trans*-azobenzene. This kind of a two-site GB potential is not a novel approach; it has been previously used in Monte Carlo simulations to study phase structure of bent-core particles with a bending angle of 140° [51] and $110 - 180°$ [52].

The orientation of the 2GB particle is parameterised by two unit vectors, $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$.

The position vector $\boldsymbol{r}$ now points to the midpoint between the two sites instead of a site centre, as in the case of the single-site GB potential, and this has to be taken into account when calculating the interaction energy. A visualisation for the 2GB model is presented in Fig. 4.4. Interaction energy between two 2GB particles, i and j, is calculated as a sum of four single-site Gay–Berne interactions:

$$\begin{aligned} &U_{\mathrm{2GB}}(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{v}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{v}}_j, \hat{\boldsymbol{r}}_{ij}) = \\ &U_{\mathrm{GB}}(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{\hat{\boldsymbol{u}}_i,\hat{\boldsymbol{u}}_j}) + U_{\mathrm{GB}}(\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{v}}_j, \hat{\boldsymbol{r}}_{\hat{\boldsymbol{u}}_i,\hat{\boldsymbol{v}}_j}) + \\ &U_{\mathrm{GB}}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{r}}_{\hat{\boldsymbol{v}}_i,\hat{\boldsymbol{u}}_j}) + U_{\mathrm{GB}}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_j, \hat{\boldsymbol{r}}_{\hat{\boldsymbol{v}}_i,\hat{\boldsymbol{v}}_j}). \end{aligned} \tag{4.10}$$

The displacement vectors $\boldsymbol{r}$ between GB interaction sites are

$$\boldsymbol{r}_{\hat{\boldsymbol{u}}_i,\hat{\boldsymbol{u}}_j} = \boldsymbol{r}_{ij} + \frac{\kappa\sigma_0}{2}(\hat{\boldsymbol{u}}_j - \hat{\boldsymbol{u}}_i), \tag{4.11}$$

$$\boldsymbol{r}_{\hat{\boldsymbol{u}}_i,\hat{\boldsymbol{v}}_j} = \boldsymbol{r}_{ij} + \frac{\kappa\sigma_0}{2}(\hat{\boldsymbol{v}}_j - \hat{\boldsymbol{u}}_i), \tag{4.12}$$

$$\boldsymbol{r}_{\hat{\boldsymbol{v}}_i,\hat{\boldsymbol{u}}_j} = \boldsymbol{r}_{ij} + \frac{\kappa\sigma_0}{2}(\hat{\boldsymbol{u}}_j - \hat{\boldsymbol{v}}_i) \text{ and} \tag{4.13}$$

$$\boldsymbol{r}_{\hat{\boldsymbol{v}}_i,\hat{\boldsymbol{v}}_j} = \boldsymbol{r}_{ij} + \frac{\kappa\sigma_0}{2}(\hat{\boldsymbol{v}}_j - \hat{\boldsymbol{v}}_i), \tag{4.14}$$

and the respective unit vectors present in Eq. (4.10) are acquired with $\hat{\boldsymbol{r}} = \boldsymbol{r}/|\boldsymbol{r}|$.

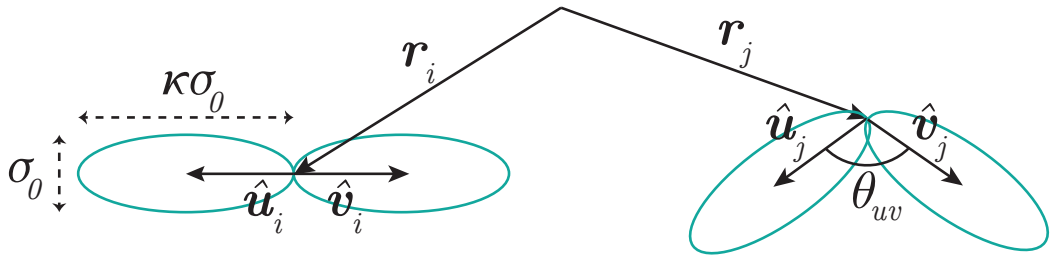


Figure 4.4: The two-site Gay–Berne models for *trans*- and *cis*-azobenzene. The models are parameterised by three vectors: the position vector $\boldsymbol{r}$ and the orientation unit vectors $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$. θ_{uv} is the bending angle between $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$. The σ_0 parameter denotes the width of a single-site Gay–Berne particle, and σ_0 multiplied by κ is the length.

4.2.4 Reduced units

When presenting thermodynamic quantities acquired from the simulation, a set of *reduced units* is used. They are marked with * and are defined as follows [31]:

Energy	$E^* = E/\epsilon_0$	(4.15)
Length	$l^* = l/\sigma_0$	(4.16)
Temperature	$T^* = k_B T/\epsilon_0$	(4.17)
Pressure	$P^* = P\sigma_0^3/\epsilon_0$	(4.18)
Volume	$V^* = V/\sigma_0^3$	(4.19)

These equations define the reduced units for the single-site Gay–Berne potential. To consider two-site Gay–Berne reduced units, parameters ϵ_0 and σ_0 have to be replaced with ϵ_0' and σ_0', the effective strength and range parameters for the two-site potential. The parameters ϵ_0 and σ_0 define the Gay–Berne units for potential distance and energy with the X configuration, as mentioned in Sec. 4.2.3. However, it is not trivial to define similar units for the two-site Gay–Berne potentials: for a bent-core molecule like *cis*-azobenzene, an X configuration cannot really be constructed. In its stead, the effective strength and range parameters ϵ' and σ' are defined with configurations where the two-site particles are positioned side by side. For *trans*-azobenzene, this is the side-by-side (SBS) configuration, and for *cis*-azobenzene it is a similar configuration seen later in Fig. 6.5b. The actual numerical values of ϵ' and σ' were acquired from the dimer interaction energy graphs, and are later presented in Table 7.1.

4.2.5 R-squared measure for goodness of fit

Coarse-graining is carried out in practice by fitting the two-site Gay–Berne model to N data points of quantum-chemical azobenzene interaction energies with a least squares algorithm. There are many statistical methods to estimate *goodness of fit*. In this study, regression analysis and more exactly, the R-squared measure, is used [53]. Data from the 2GB interactions, $\boldsymbol{y} = [y_1, y_2, \ldots y_N]^T$, are the *response variables*, which are compared to "experimental" QC data $\boldsymbol{x} = [x_1, x_2, \ldots x_N]^T$, or the *predictor variables*. Values y_i are plotted against corresponding values x_i and a least squares regression line is fitted into the dataset. Now the values y_i each have an associated *predicted value*

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \tag{4.20}$$

Here, the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ are *regression coefficients*. $\hat{\beta}_0$ is the *intercept*: it tells the expectation value of y_i when x_i equals zero. $\hat{\beta}_1$ is the *slope* that tells how much the expectation value of y_i changes as the value of x_i changes. If the model $\boldsymbol{y}$ matches

the experimental data $\boldsymbol{x}$, the coefficient values $\hat{\beta}_0 = 0$ and $\hat{\beta}_1 = 1$.

The difference between the predicted values and the response variables are the *residuals*

$$e_i = y_i - \hat{y}_i. \tag{4.21}$$

For N data points, mean of the observed data equals

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{n} y_i. \tag{4.22}$$

Now, we can estimate discrepancies between $\boldsymbol{y}$ and $\boldsymbol{x}$ by first constructing the following three sum of squares formulas: The total sum of squares

$$SSY = \sum_i (y_i - \bar{y})^2, \tag{4.23}$$

the sum of squares due to regression, or the explained sum of squares

$$SSR = \sum_i (\hat{y}_i - \bar{y})^2 \tag{4.24}$$

and the residual sum of squares, or the sum of squares due to error

$$SSE = \sum_i e_i^2 = \sum_i (y_i - \hat{y}_i)^2. \tag{4.25}$$

Note that $SSY = SSE + SSR$. Now, the fraction of the regression sum of squares and the total sum of squares is known as the *coefficient of determination* [53]

$$R^2 = \frac{SSR}{SSY} = 1 - \frac{SSE}{SSY}. \tag{4.26}$$

The value of R^2 tells to what extent the variance of $\boldsymbol{y}$ can be "explained" with $\boldsymbol{x}$. When error vanishes, the value of R^2 approaches unity.

Chapter 5

Monte Carlo simulations for an isothermal-isobaric ensemble

5.1 Computer simulation methods for molecular modelling

Computer simulations are a powerful tool for modelling complex naturally occurring phenomena and systems that cannot be solved analytically. There are two main computational methods for simulating large molecular systems [54]. Molecular dynamics (MD) is a method where the dynamic time evolution of a system of interacting particles is studied by sequentially solving its Newtonian equations of motion [55] by starting from an initial state. The MD method is deterministic: future states of a system are completely determined by its initial conditions. In the other method, Monte Carlo (MC), new configurations are generated randomly, and a predetermined criterion is applied to decide if the new configurations are accepted or not. Presence of randomness in MC means determinism is out of the picture: the state of the simulation is only linked to its direct predecessor and the probability of the state transition.

Deciding between these two techniques depends on the system ensemble and which system properties one wishes to simulate [22]. MD is well-suited for measuring time-evolution of dynamical, time-dependent variables, and MC for calculating structural quantities, e.g., building phase diagrams. This, combined with a tested simulation software, made MC a natural choice for this study. Both MD and MC can be applied to a variety of thermodynamic ensembles, but for the intents of this study, the isothermal-isobaric ensemble is examined more closely.

5.2 Classical statistical mechanics in molecular systems

In a N-particle system, the positions $\boldsymbol{r}$ and momenta $\boldsymbol{p}$ of all the particles form a $6N$-dimensional *phase space*. A point in the phase space is abbreviated as $\boldsymbol{\Gamma}$, and an *ensemble* is regarded as a collection of these points [31]. Suppose we want to simulate a system property $\mathcal{A}$ (such as energy or heat capacity) that depends on the positions and momenta of all the particles in the system. The instantaneous, fluctuating value of $\mathcal{A}$ is written as $\mathcal{A}(\boldsymbol{r}^N(t), \boldsymbol{p}^N(t)) = \mathcal{A}(\boldsymbol{\Gamma}(t))$. An experimentally observable version of $\mathcal{A}$ can be thought as its time average, which approaches the true average of $\mathcal{A}$ as the measurement time τ increases to infinity [54]:

$$\langle \mathcal{A} \rangle_{\text{time}} = \langle \mathcal{A}(\boldsymbol{\Gamma}(t)) \rangle_{\text{time}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau \mathcal{A}(\boldsymbol{\Gamma}(t)) \mathrm{d}t. \tag{5.1}$$

Calculating this becomes infeasible as system size grows larger. The *ergodic hypothesis* states that a trajectory of a point in the phase space travels through the whole ensemble in a finite time [56]. This has an important consequence: time averages of physical quantities must be the same for all members of the ensemble, or in other words, the ensemble average of the property equals its time average [54]. When the hypothesis holds, (5.1) can be replaced with

$$\langle \mathcal{A} \rangle_{\text{ens}} = \int \mathrm{d}\boldsymbol{p}^N \int \mathrm{d}\boldsymbol{r}^N \mathcal{A}(\boldsymbol{r}^N, \boldsymbol{p}^N) \rho(\boldsymbol{r}^N, \boldsymbol{p}^N) = \int \mathrm{d}\boldsymbol{\Gamma} \mathcal{A}(\boldsymbol{\Gamma}) \rho(\boldsymbol{\Gamma}), \tag{5.2}$$

where $\rho(\boldsymbol{\Gamma})$ is the *probability density* of the point $\boldsymbol{\Gamma}$ in the phase space.

In a canonical (constant-NVT[1]) ensemble, the probability density is

$$\rho_{NVT}(\boldsymbol{\Gamma}) = \frac{1}{Q_{NVT}} \exp(-\beta \mathcal{H}(\boldsymbol{\Gamma})), \tag{5.3}$$

or the *Boltzmann distribution*. Here $\beta = 1/k_B T$, and k_B is the Boltzmann constant. $\mathcal{H}$ is the system Hamiltonian resulting in the system energy E, which in turn is the sum of kinetic and potential energies: $E(\boldsymbol{r}^N, \boldsymbol{p}^N) = \mathcal{K}(\boldsymbol{p}^N) + \mathcal{V}(\boldsymbol{r}^N)$. For a canonical ensemble with N identical particles, the partition function is [31]

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int \mathrm{d}\boldsymbol{\Gamma} \exp(-\beta \mathcal{H}(\boldsymbol{\Gamma})), \tag{5.4}$$

where h is the Planck's constant. The two terms outside the integral are used to render Q_{NVT} dimensionless.

[1] Number of particles N, the system volume V and temperature T are kept constant.

The probability density of the isothermal-isobaric (constant-NPT[2]) ensemble is

$$\rho_{NPT}(\mathbf{\Gamma}) = \frac{1}{Q_{NPT}} \exp(-\beta(\mathcal{H} + PV)). \tag{5.5}$$

In constant-NPT, an additional coordinate, the now-nonconstant system volume V, is needed to comprise the phase space point $\mathbf{\Gamma}$. Thermodynamic enthalpy equals the exponent term averaged: $H = \langle \mathcal{H} \rangle + P \langle V \rangle$. The partition function is [31]

$$Q_{NPT} = \frac{1}{N!} \frac{1}{h^{3N}} \frac{1}{V_0} \int \mathrm{d}V \int \mathrm{d}\mathbf{\Gamma} \exp(-\beta(\mathcal{H}(\mathbf{\Gamma}) + PV)), \tag{5.6}$$

where the additional term V_0, the initial system volume, is needed for normalisation.

5.3 Markov chain Monte Carlo

Randomness-based methods rely on the condition that the simulation algorithm should be able to generate random states in a manner that all possible system states occur with their corresponding probabilities. This is solved by constructing a Markov chain of states [57], i.e., a sequence of trial state transitions. Using it ensures that every trial has a finite number of outcomes (a finite state space), and these outcomes only depend on initial states. An important property of Markov chain Monte Carlo (MCMC) is ergodicity [58]: any system state can be generated from any initial state with finite transitions.

Suppose we start with an initial state m with probability ρ_m. The transition to state n is represented by equation

$$\rho_n = \rho_m \pi_{mn}, \tag{5.7}$$

where π_{mn} is the probability for the transition. For N states, this equation can be replaced with a matrix equation, where the elements of the transition matrix $\boldsymbol{\pi}$ are the transition probabilities:

$$\boldsymbol{\rho}^{(N)} = \boldsymbol{\pi}^{(N)} \boldsymbol{\rho}^{(1)}. \tag{5.8}$$

The limiting distribution, when $N \to \infty$, is then

$$\boldsymbol{\rho} = \lim_{N \to \infty} \boldsymbol{\pi}^{(N)} \boldsymbol{\rho}^{(1)}. \tag{5.9}$$

The transition must satisfy the condition $\boldsymbol{\rho}\boldsymbol{\pi} = \boldsymbol{\rho}$; in other words, transitioning should not alter the equilibrium distribution $\boldsymbol{\rho}$. A simple way to meet this condition is to

[2]Instead of the system volume, pressure P is kept constant.

ensure that the stronger condition of 'microscopic reversibility' holds [31]:

$$\rho_m \pi_{mn} = \rho_n \pi_{nm}. \tag{5.10}$$

Additionally, the matrix $\boldsymbol{\pi}$ should be *stochastic*: sum of its elements should equal to one [31].

5.4 Metropolis-Hastings algorithm

In practice, given a finite length of the simulation, the transition matrix should be cleverly chosen to sample the subset of the phase space that has the greatest impact on system averages. The simplest "brute-force" Monte Carlo technique relies on evenly distributed random sampling: if we want to calculate an average of a system property $\mathcal{A}$, we just choose a large number of states m at random, calculate the value of $\mathcal{A}_m$ and then their average. In many systems, however, it is the case that some states have a greater impact on the average than others — in thermodynamic terms, their Boltzmann factor $\exp(-\beta E_m)$ is larger. To take this into account (and simultaneously lighten the calculation), evenly distributed random sampling is replaced with *importance sampling*. This is a key consequence of the Metropolis-Hastings algorithm proposed by Metropolis *et al.* [59], where configurations are chosen with a probability of their corresponding Boltzmann factor and then weighted evenly. This results in an overall faster convergence of the Markov chain [31].

In its most basic form, the Metropolis-Hastings algorithm produces a *canonical ensemble* (constant-NVT). Its trial move is executed as follows [54]:

1. Produce a new configuration by performing a random displacement of a particle. (This can also include rotation, and it can be done to several particles instead of just one.)

2. Calculate particle interactions. Based on the system potential energy $\mathcal{V}$, make a decision:

   i. If $\mathcal{V}_{\text{new}} < \mathcal{V}_{\text{old}}$, accept the new configuration.

   ii. Otherwise, generate an uniformly distributed random number between 0 and 1; $\text{ran}(0,1)$. If

   $$\exp\left(-\beta[\mathcal{V}_{\text{new}} - \mathcal{V}_{\text{old}}]\right) > \text{ran}(0,1), \tag{5.11}$$

   accept the new configuration. Otherwise, reject it.

These trial moves are executed for all the particles in a sequence — this is called a "sweep". In the Metropolis-Hastings scheme, the *acceptance ratio*, the percentage of

accepted trials, depends on the amplitude of trial move. Also, it is noteworthy that this algorithm makes transitions to higher potential energy configurations possible, based on random chance. The smaller the energy difference, the greater the probability that the transition is accepted.

5.5 Isothermal-isobaric Monte Carlo

Real-world measurements are usually performed in conditions with constant pressure and temperature. Furthermore, in a constant pressure a system is free to transform into a state where its Gibbs free energy is at a minimum, which makes first-order phase transitions possible [60]. These preambles lead us to using an isothermal-isobaric, or constant-NPT, ensemble in the simulations that follow.

For a constant-NPT ensemble, the configurational average for a system property $\mathcal{A}$ (see Eq. (5.2)), is calculated with [31]

$$\langle \mathcal{A} \rangle_{NPT} = \frac{\int_0^\infty \mathrm{d}V \exp\left(-\beta PV\right) V^N \int \mathrm{d}\boldsymbol{s} \mathcal{A}(\boldsymbol{s}) \exp\left(-\beta \mathcal{V}(\boldsymbol{s})\right)}{Z_{NPT}} \tag{5.12}$$

Here, the simulation box is a unit cube, whose volume $V = L^3$. Here, we use a set of scaled coordinates $\boldsymbol{s} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N) = L^{(-1)}\boldsymbol{r}$. This theory can be extended to include non-cubic boxes featured in the simulations of this study by replacing L with individual box dimensions: $V = X \cdot Y \cdot Z$. The configuration integral of the ensemble Z_{NPT}, or the integral over unnormalised system state probabilities, is defined by [31]

$$Z_{NPT} = \int \mathrm{d}V \exp\left(-\beta PV\right) \int \mathrm{d}\boldsymbol{r} \exp\left(-\beta \mathcal{V}(\boldsymbol{r})\right). \tag{5.13}$$

A trial configuration is conventionally generated with rules [61]

$$\boldsymbol{s}_i \to \boldsymbol{s}_i + \lambda \boldsymbol{R}^s \text{ and} \tag{5.14}$$

$$L \to L + \mu R^L, \tag{5.15}$$

where quantities R_x^s, R_y^s, R_z^s, and R^L are uniformly chosen random variables in range $(-1, +1)$; λ and μ are the displacement and volume change parameters, respectively. To achieve constant-NPT conditions, Eq. (5.15) introduces a trial volume scaling to the Monte Carlo algorithm, acting as a barostat. In practice, a volume transformation is attempted every sweep after the particle translations in Eq. (5.14). Implementation in the used simulation code differs considerably from Eq. (5.15), though. Instead of scaling the box length directly, the scaling parameter μ and a random variable R^V are chosen for volume [27, 62]

$$V \to V + \mu R^V. \tag{5.16}$$

This results in a scaling factor

$$\xi = (V + \mu R^V)/V, \tag{5.17}$$

which is then used to scale simulation box dimensions. In an isotropic volume update, every dimension is scaled together by a common factor $\xi^{1/3}$. In anisotropic variants, chosen dimensions are scaled by individually generated scaling factors ξ, and volume scaling acceptance is checked for those dimensions separately. Using this principle, one could also scale two dimensions together with $\xi^{1/2}$. This procedure ensures that volume scaling is a symmetric operation that does not break the condition of detailed balance (Eq. 5.10).

5.6 Simulation size

Real-life systems can have as much as $\mathcal{O}(10^{23})$ particles. Instead of doing the daunting task of simulating the whole system, a smaller system can be modelled with *periodic boundary conditions* (PBC): when a particle leaves the box on one side, it reappears on the other. This condition can be applied in one, two or three dimensions — bulk systems in particular use three-dimensional PBC. The problems of PBC include finite size effects due to small simulation size, and an inherent impossibility of fluctuations with a wavelength greater than the simulation box length [22].

The cell size should be considered to be greater than range of the interactions present in the simulation to prevent a particle of interacting with itself [54]. Range of interaction can be truncated with a *spherical cutoff radius* r_{cutoff}, where the interaction is cut to zero after a chosen radius. Calculations that include a cutoff radius can be made more efficient by using a Verlet list [63] for particle neighbours. It is a periodically updated list containing all the particles inside or close to entering the cutoff radius ($r_{\text{Verlet}} = r_{\text{cutoff}} +$ "skin"). Every particle has their own respective Verlet list, and interactions are only calculated for particles that are included in the Verlet list. Reconstructing the list is expensive, so it should not be done too often. Furthermore, the skin surrounding the cutoff should be thick enough so that a particle cannot penetrate through it between reconstructions [31].

5.7 Equilibration

Generally, the initial configuration for MC simulations is not close to equilibrium. A varying number of MC cycles is needed to reach stability — this phase is called *equilibration*. Length of this period can be determined by tracking the evolution of some thermodynamic quantity or structural parameter. Equilibration can also be used

to adjust the amplitude of MC trial moves to achieve a desired acceptance ratio. This breaks the prerequisite for detailed balance, or Eq. (5.10). Due to this and the generally greater fluctuations in the thermodynamic quantities, configurations from this period are not included in the analysis [30, 31]. When clear trends vanish in the tracked quantities, equilibrium can be considered reached and the *production phase* begins.

5.8 Estimating simulation results

Two kinds of errors can appear in the simulation results — systematic and statistical [22, 31]. Systematic error is the error in the physical model itself. For example, this can happen when a coarse-grained model is too coarse to represent the intricacies of the more complex, real-world system, for example as an result of size-dependence or poor equilibration. On the other hand, statistical error arises when the phase space is sampled inadequately in the calculations of the ensemble averages — for example resulting from the finite length of the simulation.

Different methods to estimate statistical error are introduced in this section. Because the Metropolis algorithm generates new configurations by making small deviations to an earlier configuration, consecutive configurations are correlated. In error estimation, random variables are assumed to be uncorrelated; thus, correlated data results in false precision. There are numerous ways to estimate statistical error in averages that are calculated from correlated data, and in this study I will principally use the blocking method documented by Flyvbjerg and Petersen [64]. Other methods include Jackknife, Bootstrap, and cross-validation [65].

5.8.1 Correlation between configurations

Suppose a quantity x is calculated in every configuration. To estimate its expectation value $\langle x \rangle$, system is assumed to be ergodic and $\langle x \rangle$ is approximated as a run average over τ_{run} configurations, or

$$m = \langle x \rangle_{run} = \bar{x} = \frac{1}{\tau_{run}} \sum_{i=1}^{\tau_{run}} x_i. \tag{5.18}$$

This finite average is a fluctuating, random variable, so its variance can be written as

$$\sigma^2(m) = \langle m^2 \rangle - \langle m \rangle^2, \tag{5.19}$$

The distance between configurations i and j, $t = i - j$, can be considered as passed simulation time. We can now write $\sigma^2(m)$ in terms of i and j: by inserting (5.18) to

Eq. (5.19), we get

$$\sigma^2(m) = \frac{1}{\tau_{run}^2}\left(\sum_{i=1}^{\tau_{run}}\gamma_{i,i} + 2\sum_{i=1}^{\tau_{run}-1}\sum_{i<j\leq\tau_{run}}\gamma_{i,j}\right), \tag{5.20}$$

where the shorthand notation

$$\gamma_{i,j} = \langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle \tag{5.21}$$

is the *covariance* of x_i and x_j [66]. In Eq. (5.20), the second term vanishes in uncorrelated data. By denoting $t = |j - i|$ and subsequently $\gamma_{i,j} = \gamma_{i,i+t}$, Flyvbjerg and Petersen write variance as

$$\sigma^2(m) = \frac{1}{\tau_{run}}\left[\gamma_0 + 2\sum_{t=1}^{\tau_{run}-1}\left(1 - \frac{t}{\tau_{run}}\right)\gamma_t\right]. \tag{5.22}$$

5.8.2 Blocking method

The blocking method relies on the assumption that correlation vanishes with sufficiently large values of t. At its core we have the following transformation. By generating a new data point as an average of two consecutive data points, the length of the dataset is halved:

$$x_i' = \frac{1}{2}\left(x_{2i-1} + x_{2i}\right), \tag{5.23}$$

$$\tau_{run}' = \frac{\tau_b}{2}. \tag{5.24}$$

In this blocking transformation, m remains invariant. Before the transformation, it is checked if τ_b is odd, and if so, one data point is left out. Subsequent blocking transformations make the data less and less correlated as the first term in Eq. (5.22) increases and the second term decreases. Flyvbjerg and Petersen state that a lower limit to variance is given by

$$\sigma^2(m) \geq \left\langle \frac{c_0}{\tau_{run} - 1}\right\rangle, \tag{5.25}$$

where

$$c_0 = \frac{1}{\tau_{run}}\sum_{k=1}^{\tau_{run}}(x_k - \bar{x})^2. \tag{5.26}$$

This can be estimated further with

$$\sigma^2(m) \approx \frac{c_0}{\tau_{run} - 1} \pm \sqrt{\frac{2}{\tau_{run} - 1}}\frac{c_0}{\tau_{run} - 1}, \tag{5.27}$$

where the second term denotes the standard error for variance.

The error estimation procedure consists of subsequent blocking transformations where variance is calculated with Eq. (5.27) after every step. Data can be considered to be uncorrelated when the variance reaches a "plateau", or the variance of a transformed dataset is within the error margins of the variance of an untransformed dataset, and vice versa. The average of x is now

$$m \pm \sigma(m) = m \pm \sqrt{\frac{c_0}{\tau_{run} - 1}}. \tag{5.28}$$

5.8.3 Statistical inefficiency

Another way to estimate correlation between subsequent blocking transformations is to measure *statistical inefficiency* [31]. It is defined as

$$s = \lim_{\tau_b \to \infty} s_b = \lim_{\tau_b \to \infty} \frac{\tau_b \sigma^2(\bar{x}_b)}{\sigma^2(x)}, \tag{5.29}$$

Here, $\bar{x}_b$ is the mean of x in a block with a size of τ_b. $\sigma^2(\bar{x}_b)$ is the variance of block averages and $\sigma^2(x)$ the variance of x in general. s_b can be plotted as a function of τ_b, and if it reaches a plateau, this maximum value of s_b can be used to approximate s. Now, s gives an estimate for the distance between uncorrelated data points x_i and x_j. Additionally, s can be used for error estimation: for uncorrelated variables, error of average of m is estimated with

$$\sigma(m) = \frac{\sigma(x_i)}{\sqrt{n/s}} \tag{5.30}$$

where n is the number of uncorrelated variables x_i. To ensure that data is uncorrelated, simulation length should be much larger than s. As a further remark, system configurations written to disk should be as uncorrelated as possible, so the value of s can also be used to estimate the frequency of disk writing.

Chapter 6

Fitting a coarse-grained model to atomistic azobenzene data

6.1 Model goals

In this chapter I will go through the process of parameterising a coarse-grained two-site Gay–Berne model (see Section 4.2.3) by fitting the model to quantum-chemical data. This parametrisation is used in the following Chapters 7 and 8 to simulate a bulk of *trans*- and *cis*-azobenzene molecules. There were two qualities the model was desired to possess:

I. The model is coarse enough so that simulations can be carried out in a matter of days.

II. There is a single "accurate enough" parametrisation that could be used to depict both *cis*- and *trans*-azobenzene molecules by only adjusting the angle between the particles.

To reach these goals, the following procedure was loosely followed:

1. Construct atomistic models for *cis*- and *trans*-azobenzene with quantum-chemical methods explained in Chapter 2.

2. Determine which two-azobenzene dimer configurations are used for the fitting procedure.

3. Calculate interaction energies for the chosen azobenzene dimers, and if possible, subject these energies to rotational averaging.

4. Fit the two-site Gay–Berne (2GB) potential to quantum-chemical data and acquire a parametrisation to be used in the simulations.

6.2 Atomistic model

6.2.1 Quantum-chemical methods

The models of *trans*- and *cis*-azobenzene were geometry optimised at the SCS-MP2 and def2-TZVPP basis set level on the TURBOMOLE software [67]. The SCS-MP2 electronic structure calculation utilises the resolution-of-identity second-order Møller-Plesset perturbation theory (RI-MP2) with spin component scaling (SCS). SCS was applied with TURBOMOLE's default values of 6/5 and 1/3 proposed by S. Grimme [12] for the same-spin and opposite-spin components, respectively. TURBOMOLE's `RICC2` module was used in the calculation. In this process, ground state energies E_A for a single *trans*- and *cis*-azobenzene molecule in vacuum were acquired. These energies and some relevant geometry parameters are presented in Table 6.1; a 3D visualisation of the geometries is seen in Fig. 4.1b.

Table 6.1: Energies and geometry parameters for *trans*- and *cis*-azobenzene. The unit vectors $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ denote the Gay–Berne site orientations and θ_{uv} is the bending angle between the vectors. For *cis*-azobenzene, these parameters are calculated from the benzene ring orientations in the geometry-optimised model, and *trans*-azobenzene is considered to be perfectly rodlike. The energy difference between the two conformations is 0.475 eV.

	$E_\text{SCS-MP2}$ [eV]	$\hat{\boldsymbol{u}}$ [Å]	$\hat{\boldsymbol{v}}$ [Å]	θ_{uv}
TAB	-15560.081	$\hat{\boldsymbol{i}}$	$-\hat{\boldsymbol{i}}$	180.0°
CAB	-15559.606	$0.8116\hat{\boldsymbol{j}} + 0.5842\hat{\boldsymbol{k}}$	$0.8116\hat{\boldsymbol{j}} - 0.5842\hat{\boldsymbol{k}}$	71.4936°

6.2.2 Dimer interaction energies and rotational averaging

The energies E_AB for two-molecule *trans-trans* and *cis-cis* dimer configurations were calculated with the same level of theory as the geometry optimisation. Multiple dimer configurations were constructed to obtain a rotationally averaged model, reflecting the rotational symmetry of the GB potential.

To distinguish the configurations, notation $(\boldsymbol{r}, \phi_A, \phi_B)$ is used. Here, $\boldsymbol{r}$ is the position vector from molecule A to molecule B, and ϕ_A and ϕ_B are their respective rotation angles along their longitudinal axes. The position vectors use the midpoint of the N-N bond as a molecular origin. Intermolecular distance is defined as $r = |\boldsymbol{r}|$. The dimer interaction energy $V(\boldsymbol{r}, \phi_1, \phi_2)$ is calculated with Eq. (2.27) without taking basis-set superposition error into account. Because *trans–cis* dimers are not considered, $E_B(B) = E_A(A)$ in every dimer configuration.

Angle-dependent configuration energies were subjected to rotational, or thermal

averaging, similarly to a previous study by Lintuvuori *et al.* [68, 69]:

$$V_{\text{average}} = \frac{\int_0^{2\pi}\int_0^{2\pi} V \exp\left(-V/k_BT\right)\mathrm{d}\phi_1\mathrm{d}\phi_2}{\int_0^{2\pi}\int_0^{2\pi} \exp\left(-V/k_BT\right)\mathrm{d}\phi_1\mathrm{d}\phi_2} \tag{6.1}$$

Now, integration over every possible rotation is approximated with a summation over rotation angles ϕ_1 and ϕ_2. The accuracy of this approximation depends directly on the number of rotation angles used to generate configurations.

$$V(\boldsymbol{r}) = \frac{\sum_{\phi_1,\phi_2} V(\boldsymbol{r},\phi_1,\phi_2)\exp\left(-V(\boldsymbol{r},\phi_1,\phi_2)/k_BT\right)}{\sum_{\phi_1,\phi_2}\exp\left(-V(\boldsymbol{r},\phi_1,\phi_2)/k_BT\right)}, \tag{6.2}$$

This process produces an energy curve with no ϕ_1- or ϕ_2-dependence. Additionally, the Boltzmann factor $\exp\left(-V(\boldsymbol{r},\phi_1,\phi_2)/k_BT\right)$ de-emphasizes positive energy values: when $V(\boldsymbol{r},\phi_1,\phi_2)$ increases, the exponential term becomes negligible. This reflects how repulsive, positive energies are less probable in the phase space. A temperature value of $T = 1000\,\mathrm{K}$ was used in the thermal averaging process.

6.3 Dimer configurations

6.3.1 *Trans*-azobenzene dimer configurations

Interaction energies $E_{\text{AB}}(AB)$ for three *trans-trans* dimers were calculated: side-by-side (SBS), end-to-end (ETE) and diagonal (DIAG) configurations, utilising a translation vector angle of 90°, 0° and 45°, respectively. In every configuration, both molecules were aligned to the x-axis and are located in a common plane — see Fig. 6.1 for visualisation. The number of rotationally averaged data points is denoted by n_{conf}.

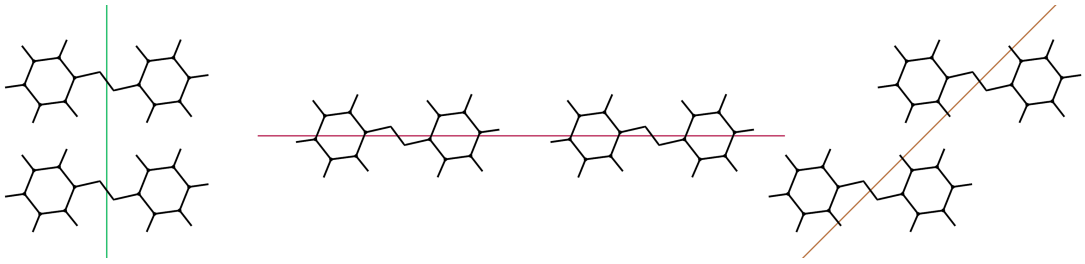


Figure 6.1: From left to right: A top-down view of the side-by-side (SBS), end-to-end (ETE) and diagonal (DIAG) *trans-trans* dimer configurations. The coloured straight lines denote the direction of translation.

6.3.1.1 Side-by-side configuration

In the SBS configurations (Fig. 6.2) the interaction energy is strongly affected by the rotation angles ϕ_1 and ϕ_2 due to short interatomic distances. Thus, a sufficiently small

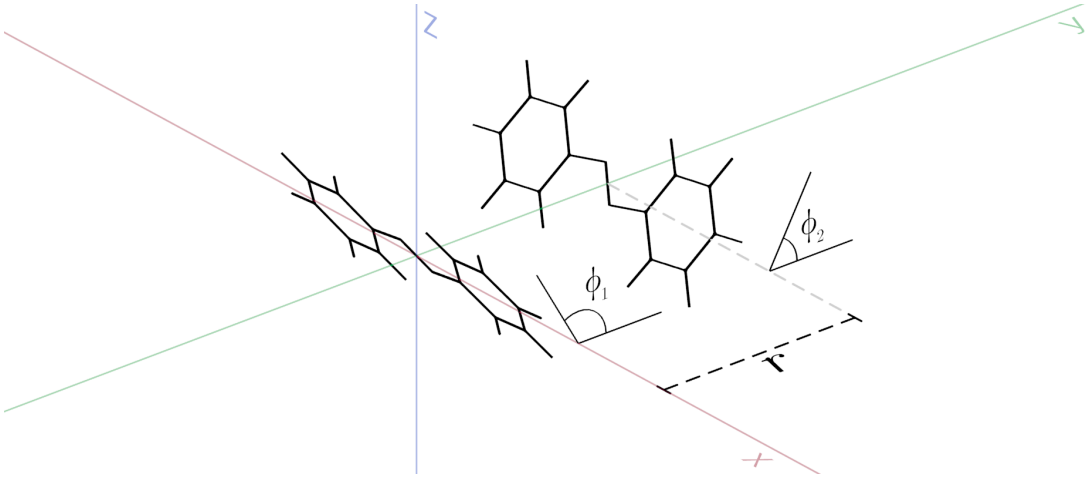


Figure 6.2: Side-by-side configuration $(\boldsymbol{r}, \phi_1, \phi_2) = (7\,\text{Å}\hat{\boldsymbol{j}}, 120°, 60°)$

rotation step of 60° had to be chosen to reflect this. A following parametrisation with ten distance values was constructed:

$$(\boldsymbol{r}, \phi_1, \phi_2) = (y\hat{\boldsymbol{j}}, m \times 60°, n \times 60°),$$
$$\text{where } m, n = 0, 1, \ldots, 5$$
$$\text{and } y = 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, \ldots, 10.0\,\text{Å } (n_{\text{conf}} = 10).$$

The rotation step of 60° gives $\left(\frac{360°}{60°}\right)^2 = 36$ configurations per constant distance. As the molecules are identical, some of the calculated orientations were redundant, e.g., $(y\hat{\boldsymbol{j}}, 0°, 180°)$ and $(y\hat{\boldsymbol{j}}, 180°, 0°)$. To reduce unnecessary energy calculations, a script was written to find which two-molecule systems were redundant. Out of the 36 configurations per constant distance, 15 could be omitted.

6.3.1.2 End-to-end configuration

In the ETE configuration (Fig. 6.3), interaction distances are much longer on average than in the SBS configuration. Thus, the rotation angles will not affect the energies as significantly, and a larger rotation step of 90° could be used. Also, due to molecules A and B sharing a common rotational axis, rotating only one molecule was necessary:

$$(\boldsymbol{r}, \phi_1, \phi_2) = (x\hat{\boldsymbol{j}}, 0°, n \times 90°),$$
$$\text{where } n = 0, 1, 2, 3.$$

Additionally, $(x\hat{\boldsymbol{j}}, 0°, 90°) \approx (x\hat{\boldsymbol{j}}, 0°, 270°)$ and $(x\hat{\boldsymbol{j}}, 0°, 0°) \approx (x\hat{\boldsymbol{j}}, 0°, 180°)$. To be exact, these geometries differ slightly, but due to long interaction distances, the effect on

energy is negligible. Therefore only two configurations were needed:

$$(\boldsymbol{r}, \phi_1, \phi_2) = (x\hat{\boldsymbol{i}}, 0°, 0°) \text{ and}$$
$$(\boldsymbol{r}, \phi_1, \phi_2) = (x\hat{\boldsymbol{i}}, 0°, 90°),$$
$$\text{where } x = 12.0, 12.5, 13.0, 13.5, 14.0, 15.0, \ldots, 18.0\,\text{Å } (n_{\text{conf}} = 9).$$

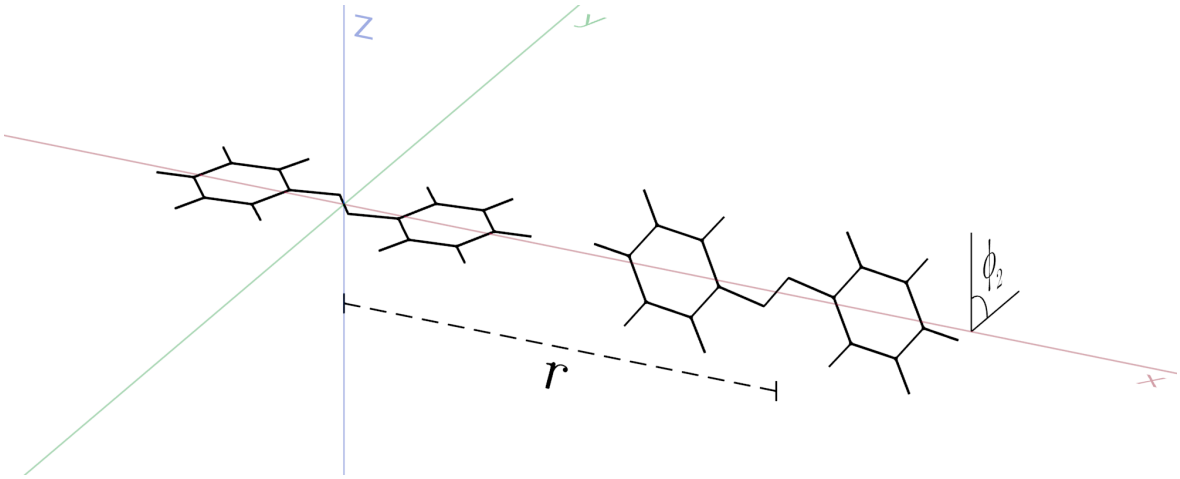


Figure 6.3: End-to-end configuration $(\boldsymbol{r}, \phi_1, \phi_2) = (13\,\text{Å}\hat{\boldsymbol{i}}, 0°, 90°)$

6.3.1.3 Diagonal configuration

The diagonal translation configuration DIAG (Fig. 6.4) is constructed with the following parametrisation, having a rotation step of 60°:

$$(\boldsymbol{r}, \phi_1, \phi_2) = (x\hat{\boldsymbol{i}} + y\hat{\boldsymbol{j}}, \phi_1, \phi_2) = (\frac{r}{\sqrt{2}}[\hat{\boldsymbol{i}} + \hat{\boldsymbol{j}}], m \times 60°, n \times 60°),$$
$$\text{where } y = x = \frac{r}{\sqrt{2}} \text{ and } m, n = 0, 1, \ldots, 5,$$
$$r = |\boldsymbol{r}| = 4.0, 4.5, 5.0, 6.0, \ldots, 10.0\,\text{Å } (n_{\text{conf}} = 7).$$

This translation scheme did not allow for omitting redundant configurations. Due to a lack of symmetry, no two orientations are interchangeable here.

6.3.2 *Cis*-azobenzene dimer configurations

For *cis*-azobenzene dimers, molecule A is fixed to an orientation where its N-N bond is aligned to the z-axis and the benzene rings face the positive y direction, where the rotation along z-axis $\phi_A = 0$. The molecule B had two orientation variants: $\phi_B = \phi_A = 0°$ and $\phi_B = 180°$, where the rings face the negative y direction. Molecule B was translated along the x-, y- and z-axes separately. Thus, there are six different

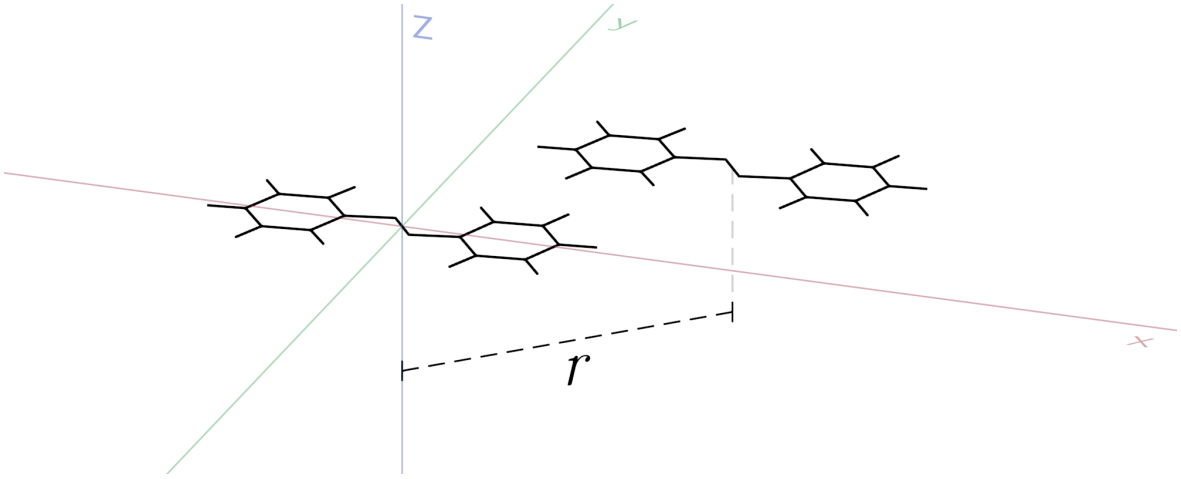


Figure 6.4: Diagonal configuration $(\boldsymbol{r}, \phi_1, \phi_2) = (\frac{10}{\sqrt{2}}\,\text{Å}[\hat{\boldsymbol{i}} + \hat{\boldsymbol{j}}], 0°, 0°)$, where $r = |\boldsymbol{r}| = 10\,\text{Å}$.

dimer configurations for CAB. In the configuration with translation along x-axis, the parametrisation is

$$
\begin{aligned}
&(\boldsymbol{r}, \phi_A, \phi_B) = (x\hat{\boldsymbol{i}} + y_0\hat{\boldsymbol{j}}, 0°, 180°),\\
&\text{where } x = 4.0, 4.5, \ldots 16.0\,\text{Å } (n_{\text{conf}} = 25) \text{ and}\\
&(\boldsymbol{r}, \phi_A, \phi_B) = (x\hat{\boldsymbol{i}}, 0°, 0°),\\
&\text{where } x = 3.5, 4.0, \ldots 12.0\,\text{Å } (n_{\text{conf}} = 18).
\end{aligned}
$$

For translation along y-axis,

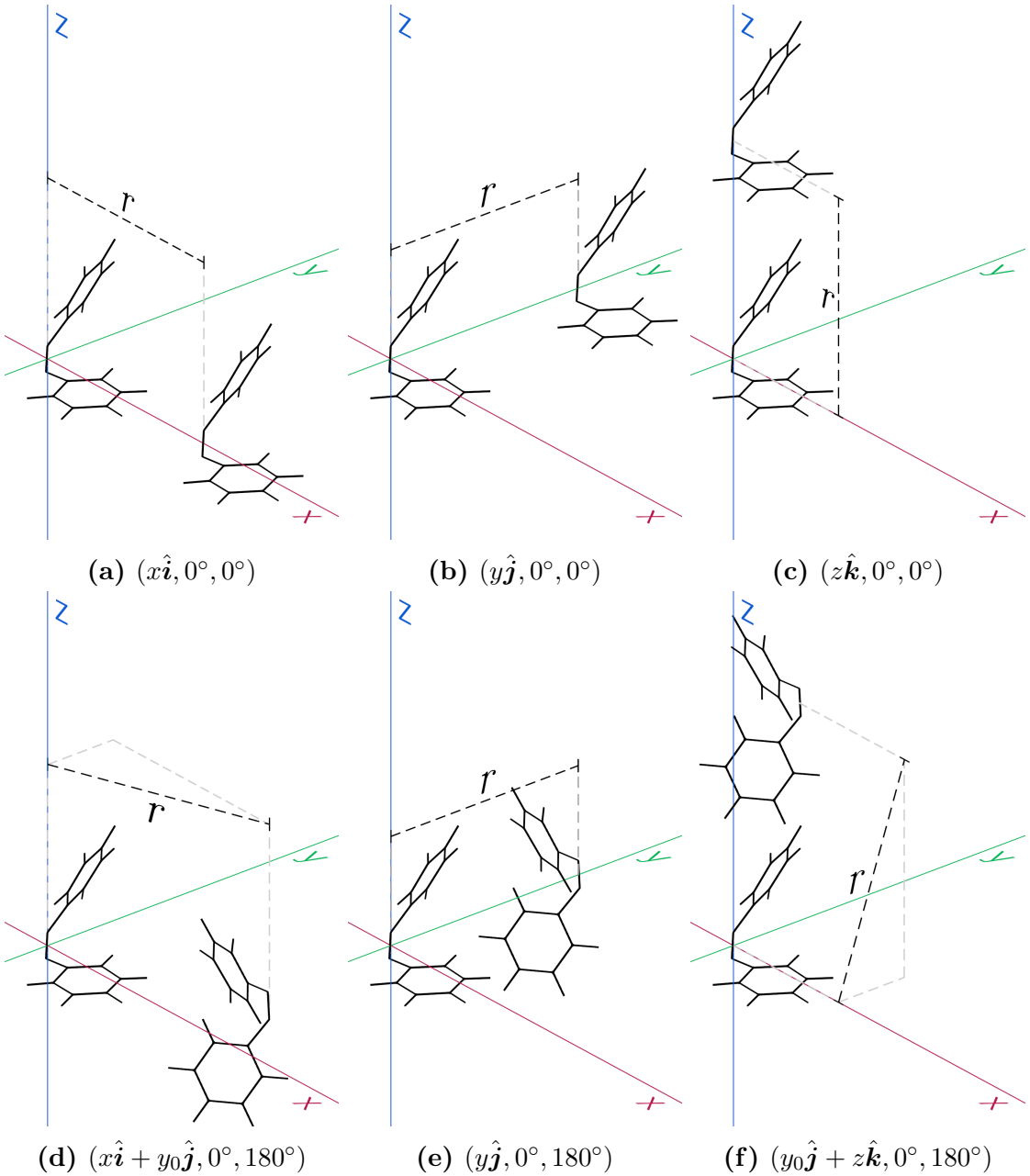
$$
\begin{aligned}
&(\boldsymbol{r}, \phi_A, \phi_B) = (y\hat{\boldsymbol{j}}, 0°, 180°),\\
&\text{where } y = 10.0, 10.5, \ldots 20.0\,\text{Å } (n_{\text{conf}} = 21) \text{ and}\\
&(\boldsymbol{r}, \phi_A, \phi_B) = (y\hat{\boldsymbol{j}}, 0°, 0°),\\
&\text{where } y = 4.5, 5.0, \ldots 10.0\,\text{Å } (n_{\text{conf}} = 12).
\end{aligned}
$$

For translation along z-axis,

$$
\begin{aligned}
&(\boldsymbol{r}, \phi_A, \phi_B) = (y_0\hat{\boldsymbol{j}} + z\hat{\boldsymbol{k}}, 0°, 180°),\\
&\text{where } z = 7.5, 8.0, \ldots 20.0\,\text{Å } (n_{\text{conf}} = 24) \text{ and}\\
&(\boldsymbol{r}, \phi_A, \phi_B) = (z\hat{\boldsymbol{k}}, 0°, 0°),\\
&\text{where } z = 8.0, 8.5, \ldots 20.0\,\text{Å } (n_{\text{conf}} = 25).
\end{aligned}
$$

In the configurations with $\phi_B = 180°$, the rotated molecule B was translated by $y_0 = 3.4944\,\text{Å}$ along the y-axis to better align the molecules. A visualisation of the configurations is presented in Fig. 6.5.

Like TAB, the CAB interaction energies were calculated with Eq. (2.27). However, rotational averaging could not be applied to the bent-core 2GB model for CAB: when rotating CAB along any axis, the orientation of the 2GB particle changes. Contrastingly, when the TAB molecule is rotated along its longitudinal axis, the 2GB particle stays intact. In principle, a rotational averaging could be used even with bent-core 2GB models by applying it separately to single-site GB particles. This was not considered in this study; for CAB, the 2GB potential was fitted directly to energies acquired from Eq. (2.27).



(a) $(x\hat{\boldsymbol{i}}, 0°, 0°)$ **(b)** $(y\hat{\boldsymbol{j}}, 0°, 0°)$ **(c)** $(z\hat{\boldsymbol{k}}, 0°, 0°)$

(d) $(x\hat{\boldsymbol{i}} + y_0\hat{\boldsymbol{j}}, 0°, 180°)$ **(e)** $(y\hat{\boldsymbol{j}}, 0°, 180°)$ **(f)** $(y_0\hat{\boldsymbol{j}} + z\hat{\boldsymbol{k}}, 0°, 180°)$

Figure 6.5: Configurations for the CAB molecules A and B. The molecule A is fixed to the origin and B is translated along a Cartesian axis by x,y or z. In the top row, both molecules are facing the same direction. In the bottom row, B is rotated 180° along the z-axis. The $(\boldsymbol{r}, \phi_A, \phi_B)$ notation is used to label the configurations.

6.4 Least squares fit

6.4.1 Methods

SciPy, the Python library for scientific computing, was used for the fitting. The least squares method `scipy.optimize.least_squares` [70] introduced in Scipy 0.17.0 was utilised, namely for its feature of parameter bounds. A three-point method was used to numerically estimate the Jacobian matrix, and the trust region reflective algorithm [71] was used for minimisation.

6.4.2 Residual weighting

In the least squares fitting process, the residual (4.21), or the difference between the descriptive model (in this case, two-site Gay–Berne) and the original (quantum-chemical) data, is calculated for each data point i:

$$e_i = V_{\mathrm{2GB}}(\boldsymbol{r})_i - V_{\mathrm{QC}}(\boldsymbol{r})_i \tag{6.3}$$

The least squares algorithm aims to minimise the sum

$$\sum_i e_i^2 = \sum_i [V_{\mathrm{2GB}}(\boldsymbol{r})_i - V_{\mathrm{QC}}(\boldsymbol{r})_i]^2. \tag{6.4}$$

It should be noted that energies with larger absolute values influence the resulting parametrisation more than smaller ones. This leads to bias particularly in the repulsive domain, where potential energy approaches infinity when r approaches zero. Additionally, this has a consequence of DIAG and SBS configurations with deeper potential wells being better represented in the TAB fit than ETE with a lower potential well. These effects can be countered with a Boltzmann factor

$$\exp\left(-\frac{V_{\mathrm{QC}}(\boldsymbol{r})_i - V_{\min}}{k_B T}\right). \tag{6.5}$$

Here, $V_{\min}$ is the lowest quantum-chemical energy of the whole dataset to which the 2GB model is being fitted into. The chosen temperature was the same as in rotational averaging (Section 6.2.2, $T = 1000\,\mathrm{K}$).

In addition, the number of data points $V(r)$ varies between different configurations. Configurations with more data points would influence the fit more, so to counter this, the residual e_i is divided by the number of data points in the configuration, or the normalising factor n_{conf}. Together, these two factors form a correction coefficient

$$N_{\mathrm{corr}} = \frac{1}{n_{\mathrm{conf}}} \exp\left(-\frac{V_{\mathrm{QC}}(\boldsymbol{r})_i - V_{\min}}{k_B T}\right). \tag{6.6}$$

Now, Eq. (6.4) becomes

$$\sum_i N_{\text{corr}}\, e_i^2 = \sum_i \frac{1}{n_{\text{conf}}} \exp\left(-\frac{V_{\text{QC}}(\boldsymbol{r})_i - V_{\text{min}}}{k_B T}\right) [V_{\text{2GB}}(\boldsymbol{r})_i - V_{\text{QC}}(\boldsymbol{r})_i]^2. \tag{6.7}$$

6.4.3 Fitting parameters

A parametrisation for the GB potential by Bates and Luckhurst with parameters (κ, κ', μ, ν) = (4.4, 20, 1, 1) is known to produce crystal (Cr), smectic-A (SmA), nematic (N) and isotropic (I) phases [29, 62, 68], and provided an ansatz for the fitting. To take into account that the two-site GB potential consists of two conjoined GB particles, the GB length-to-breadth ratio κ was divided in half to keep the length-to-breadth ratio of a whole rodlike two-site TAB particle at 4.4. Initially, all four parameters (κ, κ', μ, ν) were kept constant during the fitting, σ_0 and ϵ_0 being the only fitting parameters. It was quickly noticed that this scheme was too strict to accurately portray the *cis* and *trans* isomers, so parameters κ and κ' were included as fitting parameters as well.

6.4.4 Gay–Berne model parametrisation

When attempting to fit the 2GB model to all data points at once, some problems arose. The fitting parameters tended to reach for their respective upper or lower bounds — suggesting that the model struggled to depict both isomers at once. Thus it was decided to first fit the 2GB potential to data from the *trans* configurations only. Then, the acquired potential would be used as an ansatz for a *cis* configuration fit.

SBS and DIAG configurations dominated the *trans* isomer fit, even after applying the adjustments described in the section 6.4.2. A successful fit was only acquired when the importance of the ETE orientation was artificially weighted more. The value of the normalising factor n_{ETE} was ultimately lowered from 9 to 0.0001. The acquired parametrisation is seen in Fig. 6.6, and it could now be used as an ansatz for a combined *cis* and *trans* molecule fit. This resulted, however, in a parametrisation that could not depict properly neither of the configurations, so the idea of a single parametrisation was discarded. A separate fit was done for CAB, and the resulting energy graphs are seen in Figure 6.7. Here, no arbitrary weighting had to be done to get a reasonable parametrisation; the only configuration with slightly mismatching energies was $(y\hat{\boldsymbol{j}}, 0°, 180°)$.

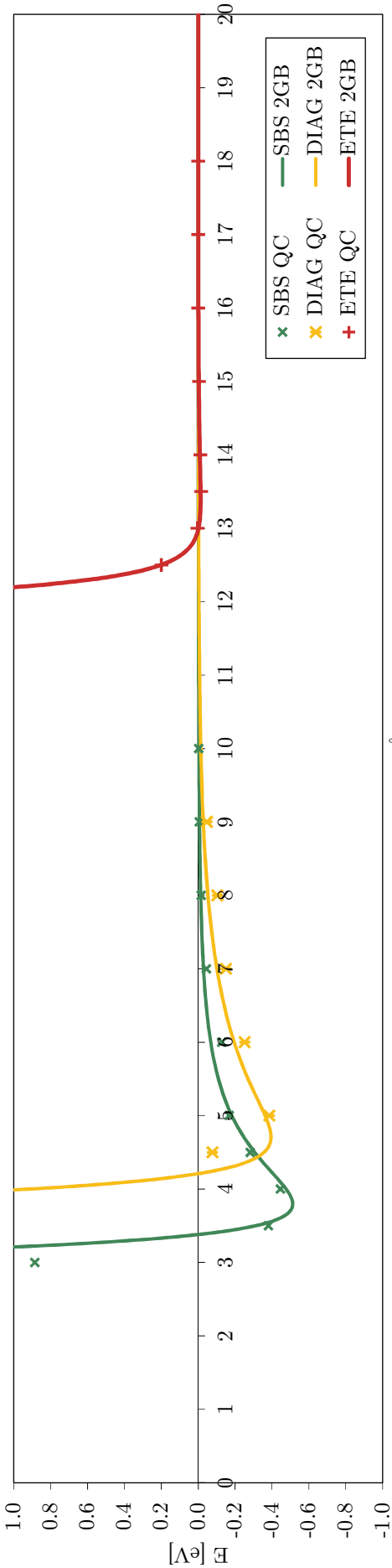
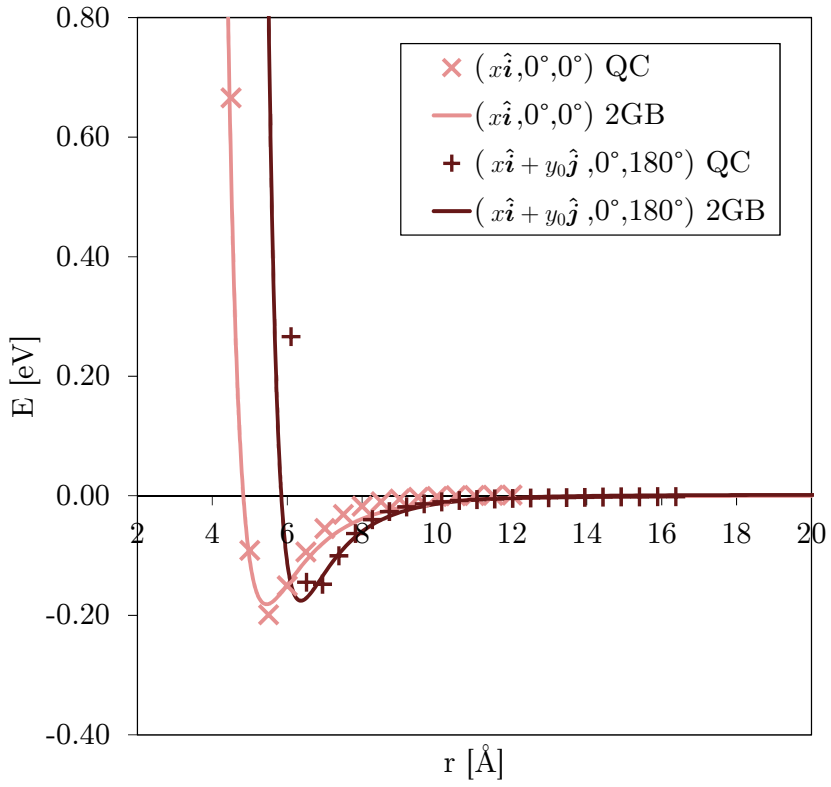
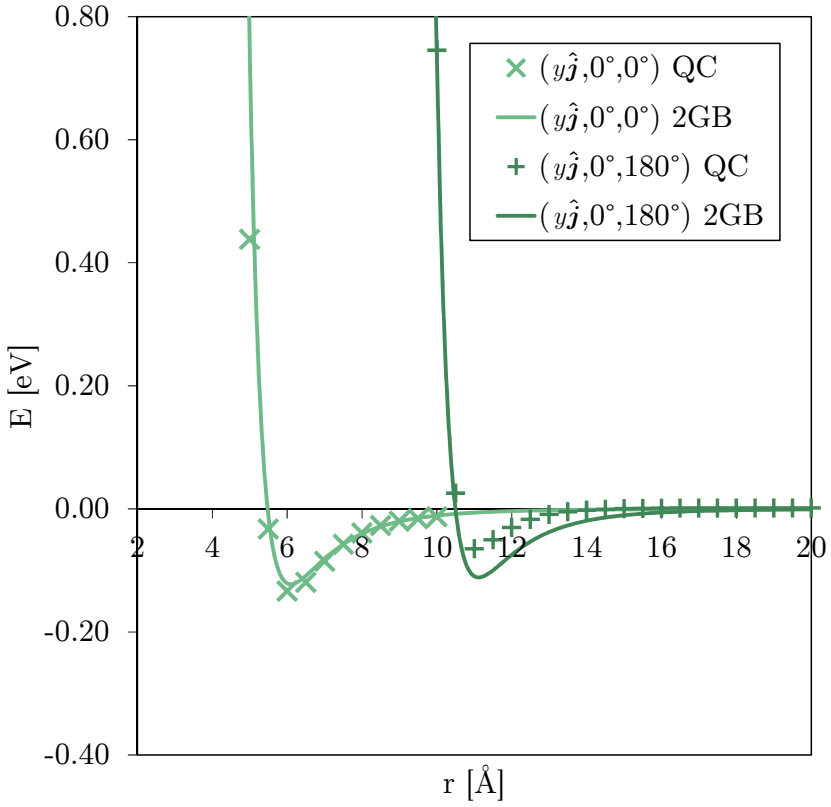


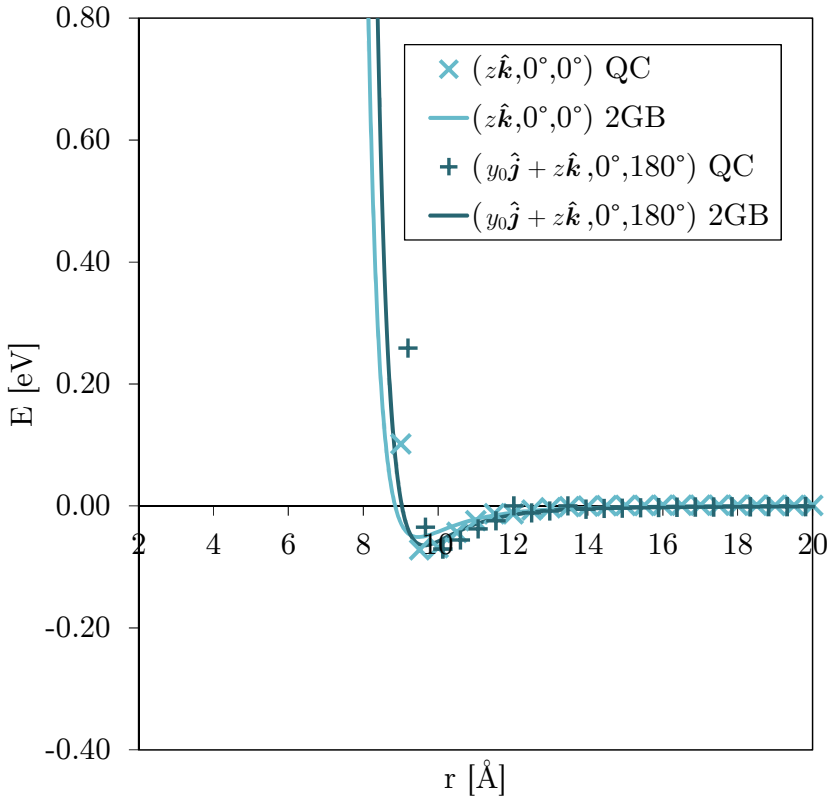
Figure 6.6: Two-site Gay–Berne potential (2GB) fitted to three configurations (SBS, ETE, DIAG) of rotationally averaged quantum-chemical (QC) interaction energies of *trans*-azobenzene dimers.



(a) *Cis*-azobenzene dimer translation along x-axis.



(b) *Cis*-azobenzene dimer translation along y-axis.



(c) *Cis*-azobenzene dimer translation along z-axis.

Figure 6.7: Two-site Gay–Berne (2GB) potential fitted to six configurations of *cis*-azobenzene dimer quantum-chemical (QC) interaction energies. The $(\boldsymbol{r}, \phi_A, \phi_B)$ notation is used to differentiate the configurations. See Fig. 6.5 for corresponding visualisations.

6.5 Error estimation

6.5.1 Error estimation for quantum-chemical energies

Counterpoise correction was used to find the magnitude of basis set superposition error (BSSE, see Section 2.3.3). This was applied to a configuration where intermolecular distance is at its minimum: the SBS configuration of TAB. Three configurations were examined, and the results are seen in Table 6.2. BSSE is quite significant when the molecules are close ($E_{\mathrm{AB}}^{\mathrm{BSSE}} = 0.4146\,\mathrm{eV}$ for $(4\hat{\boldsymbol{j}}, 0°, 60°)$), but it greatly reduces as molecules separate: when intermolecular distance increases from 4 Å to 5 Å, $E_{\mathrm{AB}}^{\mathrm{BSSE}}$ of configuration $(\boldsymbol{r}, 120°, 120°)$ drops from 0.1442 eV to 0.0415 eV. As expected, the energies without BSSE correction are lower than the corrected ones. Despite BSSE causing a visible change in energy, counterpoise correction was not applied to the calculated energies. Applying it would have been a slight overstatement, especially when taking into account the fact that much more significant details than BSSE are

already lost in rotational averaging and the fitting procedure itself.

Another reference calculation was also made — one without spin-component scaling (SCS). As seen in Table 6.3, absence of SCS provided deeper potential wells, as expected: in MP2 calculations without SCS correction, the potential wells are generally too deep, SCS-corrected energies being closer to empirical results [12].

Table 6.2: Counterpoise correction for three *trans*-azobenzene side-by-side configurations. The unit of all energy values is eV, and the unit of distance is Å. $E_{\mathrm{AB}}^{\mathrm{BSSE}} = \Delta E_{\mathrm{AB}}^{\mathrm{CP}} - \Delta E_{\mathrm{AB}}$ tells the magnitude of basis set superposition error, where ΔE_{AB} and $\Delta E_{\mathrm{AB}}^{\mathrm{CP}}$ are calculated with Equations (2.27) and (2.28), respectively.

$(\boldsymbol{r}, \phi_1, \phi_2)$	$E_{\mathrm{AB}}(\mathrm{AB})$	$E_{\mathrm{AB}}(\mathrm{A})$	$E_{\mathrm{AB}}(\mathrm{B})$	$\Delta E_{\mathrm{AB}}^{\mathrm{CP}}$	ΔE_{AB}	$E_{\mathrm{AB}}^{\mathrm{BSSE}}$
$(4\hat{\boldsymbol{j}}, 0°, 60°)$	-31115.8046	-15560.2080	-15560.3685	4.7719	4.3573	0.4146
$(4\hat{\boldsymbol{j}}, 120°, 120°)$	-31120.6090	-15560.1530	-15560.1531	-0.3029	-0.4471	0.1442
$(5\hat{\boldsymbol{j}}, 120°, 120°)$	-31120.3371	-15560.1017	-15560.1017	-0.1336	-0.1752	0.0415

Table 6.3: Effect of spin-component scaling on *trans*-azobenzene energies. The unit of energy is eV. The values of ΔE_{AB} are calculated with Eq. (2.27).

$(\boldsymbol{r}, \phi_1, \phi_2)$	$E_{\mathrm{AB}}^{\mathrm{MP2}}$	$E_{\mathrm{AB}}^{\mathrm{SCS\text{-}MP2}}$	$\Delta E_{\mathrm{AB}}^{\mathrm{MP2}}$	$\Delta E_{\mathrm{AB}}^{\mathrm{SCS\text{-}MP2}}$	ΔV
$(4\hat{\boldsymbol{j}}, 0°, 60°)$	-31118.2640	-31115.8046	1.8979	4.3573	2.4594
$(4\hat{\boldsymbol{j}}, 0°, 120°)$	-31118.6576	-31116.2276	1.5042	3.9343	2.4301
$(4\hat{\boldsymbol{j}}, 120°, 120°)$	-31122.6063	-31120.6090	-2.4444	-0.4471	1.9974
$(4\hat{\boldsymbol{j}}, 120°, 60°)$	-31121.6218	-31119.5354	-1.4599	0.6265	2.0865

6.5.2 Goodness of fit

Goodness of fit was examined with R^2 measure regression analysis. Here, nomenclature and methods explained in Section 4.2.5 are used. Energies from the acquired Gay–Berne model $\boldsymbol{y}$ were compared to quantum-chemical data $\boldsymbol{x}$ by plotting data points y_i against corresponding points x_i. Least squares regression lines of the form of Eq. (4.20) were calculated, and the coefficients of determination R^2 were determined with Eq. (4.26) (without using the correction coefficient (6.6)). The regression lines and their respective equations are presented in Fig. 6.9.

As multiple curves were fitted simultaneously, some of the fitted 2GB potentials and their respective quantum-chemical energy curves could differ considerably. Discrepancies are visible especially in the positive repulsive region (see Figures 6.6 and 6.7), e.g., in the ETE configuration, $V_{\mathrm{2GB}}(3\,\text{Å}) = 4.371\,\mathrm{eV}$ and $V_{\mathrm{QC}}(3\,\text{Å}) = 0.885\,\mathrm{eV}$. To see more clearly how closely the parametrisation depicts the attractive region of the potential (the actual area of interest in the fit), all positive energy values were omitted

from the regression analysis. Additionally, one negative energy value with a stark difference between the QC and 2GB energies was discarded: in the DIAG configuration, $V_{\mathrm{2GB}}(4.5\,\text{Å}) = -0.356\,\text{eV}$ and $V_{\mathrm{QC}}(4.5\,\text{Å}) = 0.0778\,\text{eV}$. As seen in Fig. 6.6, the QC graph rises much more steeply than the fitted 2GB graph. (In SBS, it is exactly the opposite. Thus, it might be impossible to fit the 2GB model to this dataset with all data points depicted adequately.) Some outliers are visible in Figures 6.9a and 6.9b, marked in red. In total, 5 out of 26 and 16 out of 125 data points were omitted for TAB and CAB, respectively.

After the modifications discussed above, the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ were satisfactory for both conformations — close to zero and one, respectively. Similarly, the value of R^2 was close to the desired value of one in both conformations: for TAB $R^2 = 0.9626$ and for CAB $R^2 = 0.9054$. This indicates that both parametrisations depict their respective isomers (if not some of the singular data points) reasonably well.

6.6 Resulting model

Two separate models for *trans*- and *cis*-azobenzene, seen in Fig. 6.8, were acquired. This means that out of the two objectives stated in 6.1, only the first one was reached. The reason for this can be seen when observing the parameters of the two models, which differ significantly. The largest differences manifest themselves in the contact distance ϵ_0 and the well depth ratio κ': $\epsilon_{0,\mathrm{TAB}}$ is almost four times the value of $\epsilon_{0,\mathrm{CAB}}$, and $\kappa'_{0,\mathrm{TAB}}$ is almost ten times the value of $\kappa'_{0,\mathrm{CAB}}$. It appears that a single parametrisation was not acquirable, not at least without unlocking parameters μ and ν for fitting.

Table 6.4: Gay–Berne model parameters acquired from the least squares fit. Parameters marked with an asterisk (*) were constant during the fit. Reference parameters [29] are from a single-site GB study, [51] and [52] from two-site studies.

	ϵ_0	σ_0	κ	κ'	μ^*	ν^*
Ref. [29]			4.4	20	1	1
Ref. [51]			3	5	1	2
Ref. [52]			3	5	2	1
TAB	0.201897433	3.387895922	1.916408439	18.62859264	1	1
CAB	0.055549567	4.961860259	1.224314214	2.01296207	1	1

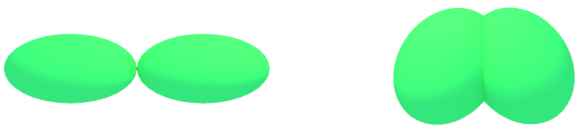
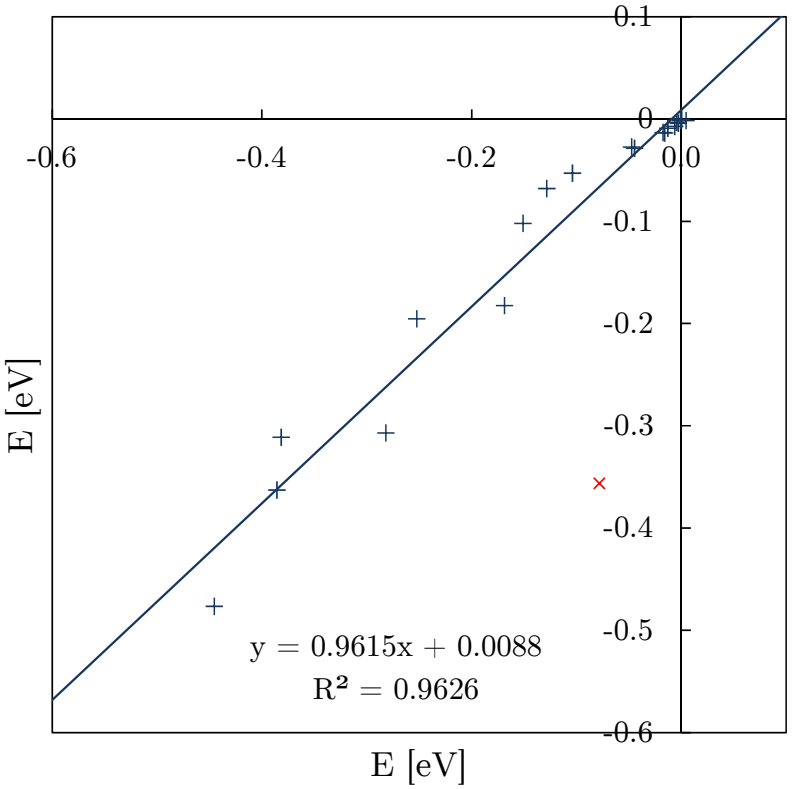
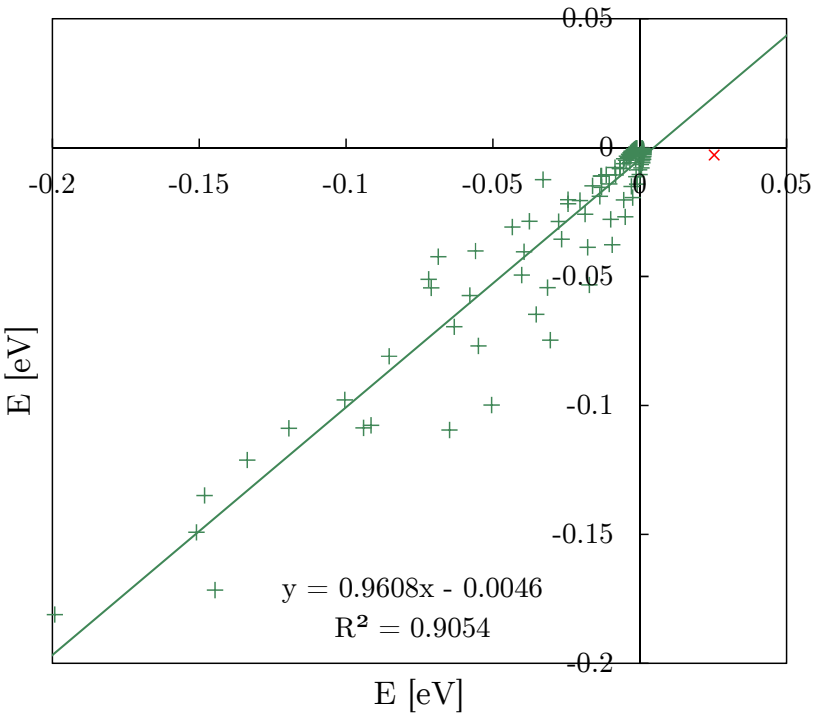


Figure 6.8: Visualisation of the acquired two-site Gay–Berne models for *trans*- and *cis*-azobenzene. Compare to atomistic models in Fig. 4.1b. The parameters are seen in Table 6.4.



(a) *Trans*-azobenzene.



(b) *Cis*-azobenzene.

Figure 6.9: Regression lines for *trans*- and *cis*-azobenzene, where the two-site Gay–Berne energies are plotted as a function of quantum-chemical energies. Some data points were disregarded due to a significant discrepancy between the fitted potential and the QC energy in an unimportant domain. The disregarded data points are marked in red (×).

Chapter 7

Simulating a bulk system of azobenzene molecules

7.1 Software

A Fortran 2008 simulation software written by Jouni Karjalainen, PTGBCYL [30], was used for the Metropolis Monte Carlo simulations. The software implements regular single-site Gay–Berne (GB) potential, Lennard-Jones (LJ) potential and a mixed GB-LJ potential, and these potentials can be studied in bulk or as confined to a cylindrical cavity. An expansion was written to the software to include support for two-site Gay–Berne (2GB) potential. PTGBCYL additionally provides a method for parallelisation with parallel tempering. Using this feature was briefly considered, but it was not ultimately applied in this study.

Despite initial interest in simulating *trans–cis*-interaction, all the simulations in this study are either all-*trans* or all-*cis* simulations. This restriction arose from the differing Gay–Berne parametrisations for the *cis* and *trans* molecules (see Table 6.4). Implementing a mixed *trans–cis* simulation would have required significant alterations to the software, because PTGBCYL treats the σ_0 and ϵ_0 parameters as unities.

7.2 Initialisation

All simulations contain $N = 2000$ 2GB particles in a cuboid simulation box with periodic boundary conditions. The used ensemble was NPT, or isothermal-isobaric. During the equilibration period (see Section 5.7), the amplitude of translation and rotation were adjusted to reach a trial move acceptance ratio of 30%, and similarly volume scaling was adjusted to reach a volume scaling acceptance ratio of 25%.

The cutoff radius for the Gay–Berne interaction had a value of $r_{cutoff} = 4\sigma_0 = 13.55\,\text{Å}$ throughout the simulations. The radius is rather cumbersomely implemented

between the 2GB dimer midpoints, instead of between individual interaction sites. Due to this, the cutoff radius is rather short for the used parametrisation. In SBS and DIAG configurations, the dimer interaction energy V_{GB} is close to zero at this radius as it should be (see Fig. 6.6), but in the ETE configuration, $V_{\mathrm{GB}} = -0.013\,\mathrm{eV}$ which is still definitely substantial: it is in fact approximately equal to the ETE potential well depth. In fact, a cutoff radius this large could possibly even affect the phase structure of the model. An interaction site-based cutoff radius would have been a more reasonable approach, and would have ensured that ETE potential graph would not have been prematurely cut off.

Verlet list (see Section 5.6) was used to optimise interaction calculations by disregarding interactions for particles whose distance from each other was larger than the Verlet radius. The value of Verlet radius r^*_{Verlet} was adjusted during the simulation by calculating the value of "skin" on top of the cutoff radius to equal at least two times the maximum translation — this is to ensure that interparticular distance would not become smaller than the cutoff radius between Verlet list reconstructions.

The GB interaction energy (4.2) increases rapidly when radius between two GB particles approaches zero. To improve numeric stability, a "hard core" for the molecules was implemented in the software where translations that resulted in a separation (4.3) R value less than 0.6 were discarded.

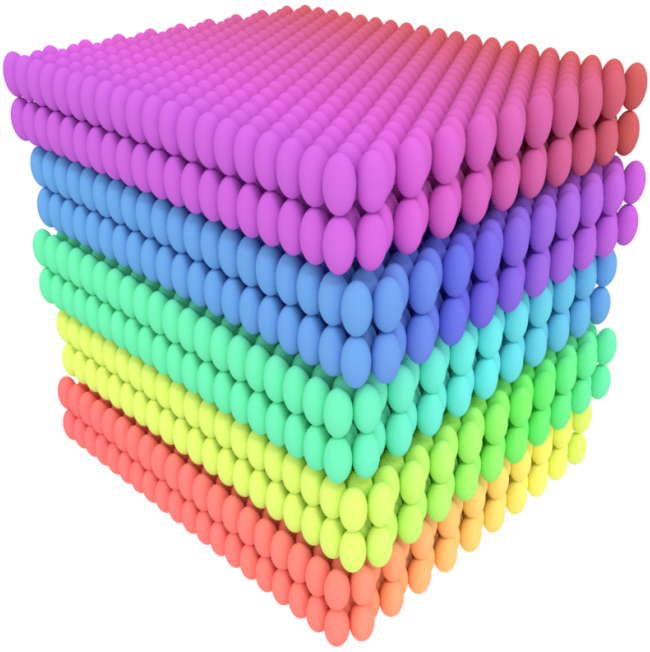
7.3 Test simulations

7.3.1 *Trans*-azobenzene

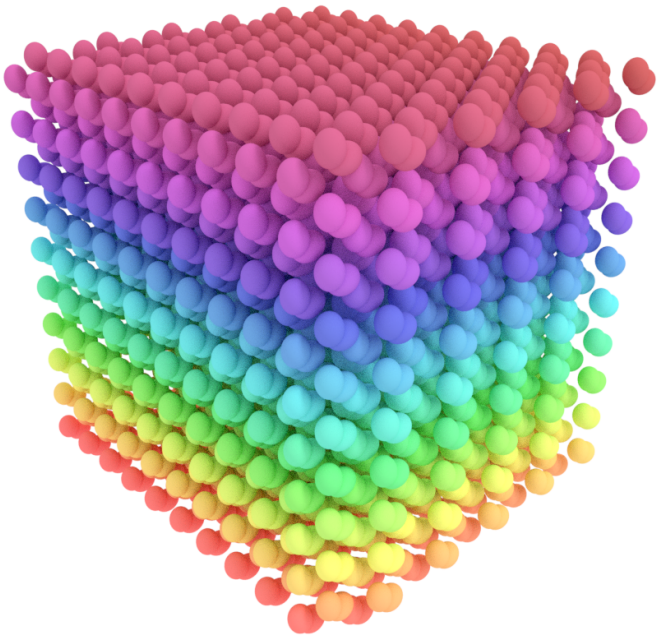
The GB potential parametrisation (4.4, 20, 1, 1) was known to produce interesting Cr–SmA–N–I phase structure [29, 62, 68], and does not differ greatly from the *trans*-azobenzene 2GB parametrisation (1.916, 18.629, 1, 1)[1]. Thus, the temperature range $T^* = 0.8 \ldots 1.9$ and pressure $P^* = 2.0$ from earlier studies were a reasonable ansatz for suitable simulation conditions. The conditions were examined by running test simulations with less temperature points than in simulations proper: pressure values $P^* = 1.0$, 2.0 and 4.0 were simulated at temperatures $T^* = 0.5, 0.75, \ldots, 2.0$.

The initial state of the *trans*-azobenzene simulations consists of z-axis-aligned particles on hexagonal close-packed layers along the xy plane, as seen in Fig. 7.1a. The ensemble was heated to the maximum temperature of $T^* = 2.00$, from where the temperature was gradually cooled down. The calculated reduced enthalpies are presented in Fig. 7.2 as a function of temperature. The steep transition between temperatures

[1]Note that because this study uses a two-site GB model and the aforementioned studies a single-site model, the value of $\kappa = 1.916$ should be compared here to 2.2 instead of 4.4, thus resulting in seemingly "similar" parametrisations. See discussion in Section 6.4.3.



(a) *Trans*-azobenzene particles in five hexagonal close-packed layers.



(b) *Cis*-azobenzene particles in twenty ten-by-ten layers.

Figure 7.1: The initial states for azobenzene simulations with 2000 two-site Gay–Berne particles. The particles are colour-coded with unique hue values to better visualise which Gay–Berne sites form a two-site particle.

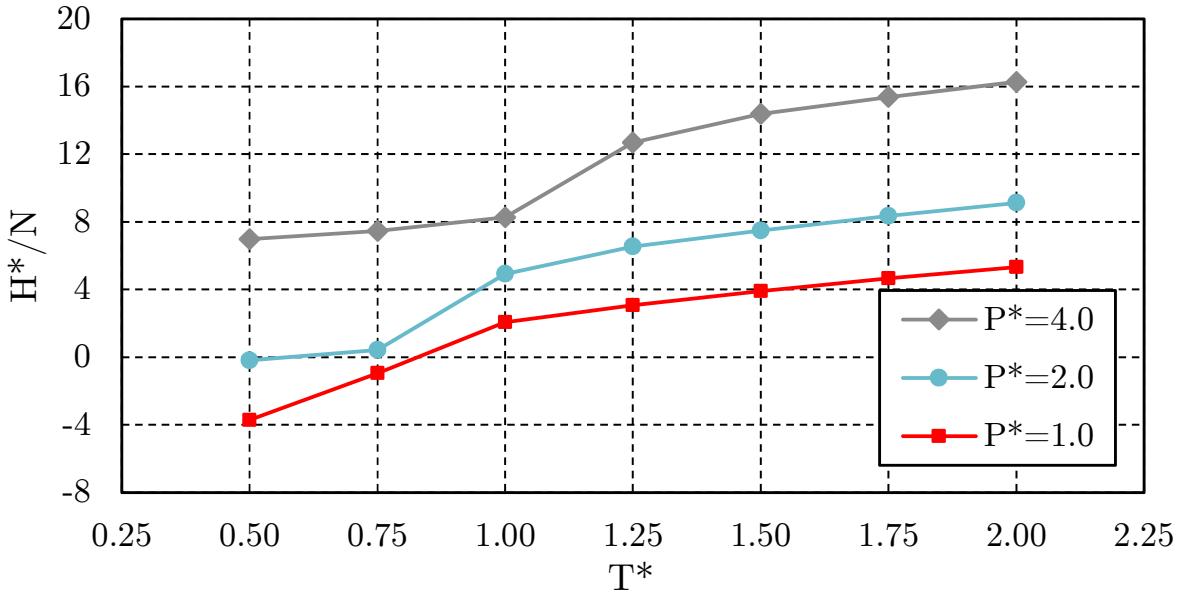


Figure 7.2: Reduced enthalpies per particle as a function of reduced temperature for the initial test simulations of *trans*-azobenzene at three pressure values.

$T^* = 1.00$ and 1.25 suggests that a pressure of $P^* = 4.0$ is high enough to support liquid crystal phases, so it was used in production simulations. Also, the tested temperature range was extended to lower temperatures: the production simulations were ultimately carried out at a temperature range of $T^* = 0.40 \ldots 2.00$, $\Delta T^* = 0.02$.

7.3.2 *Cis*-azobenzene

The *cis*-azobenzene 2GB parametrisation (1.224, 2.013, 1, 1) differs from the reference parametrisation (4.4, 20, 1,1) quite considerably. To simulate CAB in the same conditions as TAB, their absolute temperatures and pressures should be equal: $T_{cis} = T_{trans}$ and $P_{cis} = P_{trans}$. Using this equality, the reduced pressures and temperatures for CAB can be calculated with the following expressions derived from Equations (4.17) and (4.18):

$$T^*_{cis} = T^*_{trans} \frac{\epsilon'_{trans}}{\epsilon'_{cis}} \tag{7.1}$$

$$P^*_{cis} = P^*_{trans} \frac{\epsilon'_{trans}}{\epsilon'_{cis}} \left(\frac{\sigma'_{0cis}}{\sigma'_{0trans}} \right)^3 \tag{7.2}$$

By plugging the values of ϵ' and σ' from Table 7.1 into Equations (7.1) and (7.2), we get the pressures and values seen in Table 7.2, which were used as the conditions for CAB initial test simulations.

The initial state of the *cis*-azobenzene simulations is a ten-by-ten-by-twenty box where all the particles point into a common direction, see Fig. 7.1b. Unlike with TAB, the results from CAB simulations suggested that there is no phase transition present

in the simulated pressure and temperature ranges (see Fig. 7.3). Enthalpy dependence on temperature is approximately linear in the whole simulation range. Thus, full-scale CAB production simulations were not carried out.

Table 7.1: Effective strength and range parameters σ_0' and ϵ_0' for *trans*- and *cis*-azobenzene. The configurations used to determine the parameter values are side-by-side (SBS) seen in Fig. 6.6 and $(y\hat{\boldsymbol{j}}, 0°, 0°)$ seen in Fig. 6.5b, respectively.

	ϵ_0' [eV]	σ_0' [Å]
TAB	0.5122583477	3.38
CAB	0.1773223156	5.82

Table 7.2: Reduced temperatures and pressures for *trans*- and *cis*-azobenzene. For comparison, the absolute temperatures and pressures acquired with Eqs. (4.17) and (4.18), respectively, are presented. However, the realisticity of these rather astronomical values should be taken with a grain of salt.

T^*_{trans}	T [K]	T^*_{cis}
0.50	2972	1.4444
0.75	4458	2.1666
1.00	5945	2.8889
1.25	7431	3.6111
1.50	8917	4.3333
1.75	10403	5.0555
2.00	11889	5.7777

P^*_{trans}	P [Pa]	P^*_{cis}
1.00	2.13E+14	14.78
2.00	4.25E+14	29.57
4.00	8.50E+14	59.14

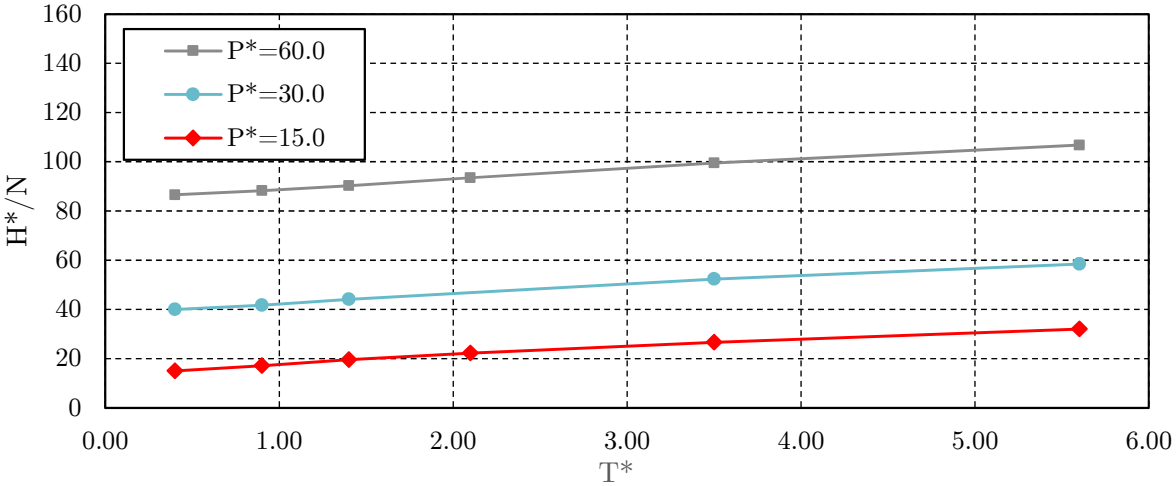


Figure 7.3: Reduced enthalpies per particle as a function of reduced temperature for the initial simulations of *cis*-azobenzene at three pressure values.

7.4 Simulation snapshots

A simulation snapshot is saved to disk every 50 Monte Carlo sweeps. It contains the position and orientation of every 2GB particle, the dimensions of the simulation box, thermodynamic quantities calculated from particle interactions and additional simulation data regarding the translation, rotation and scaling operations. The simulation box was scaled in the directions of the Cartesian axes x, y and z. At temperatures $T^* < 1.06$, all axes were scaled anisotropically (see Section 5.5), as is customary when orientational long-range order is present to acquire isotropic pressure throughout the simulation box. When comparing to initial test simulations with isotropic scaling at the same temperatures, this also aligned the director vector more closely to z-axis. Nevertheless, as seen later in Fig. 8.13, the director still quite visibly differs from z-axis. At temperatures $T^* \geq 1.06$, a.k.a. the temperature range identified as the isotropic phase, all simulation box dimensions were scaled isotropically (Eq. 5.15)).

A simulation typically started with a 100000–sweep long equilibration period, followed by at least a 100000–sweep long production period. At several temperatures, especially at those close to a phase transition, 100000 sweeps were not enough to equilibrate the system. In such cases the simulation would be extended further (see Fig. 7.4). Simulation length as a function of temperature is presented in Fig. 7.5. Quantity averages presented in the following chapter were calculated from the snapshots where the simulation was considered to be equilibrated. Equilibration was determined from the energy curves, which might result in cases where equilibrium would not be actually reached. A more rigorous approach would be to use radial density functions to determine if equilibrium has been reached.

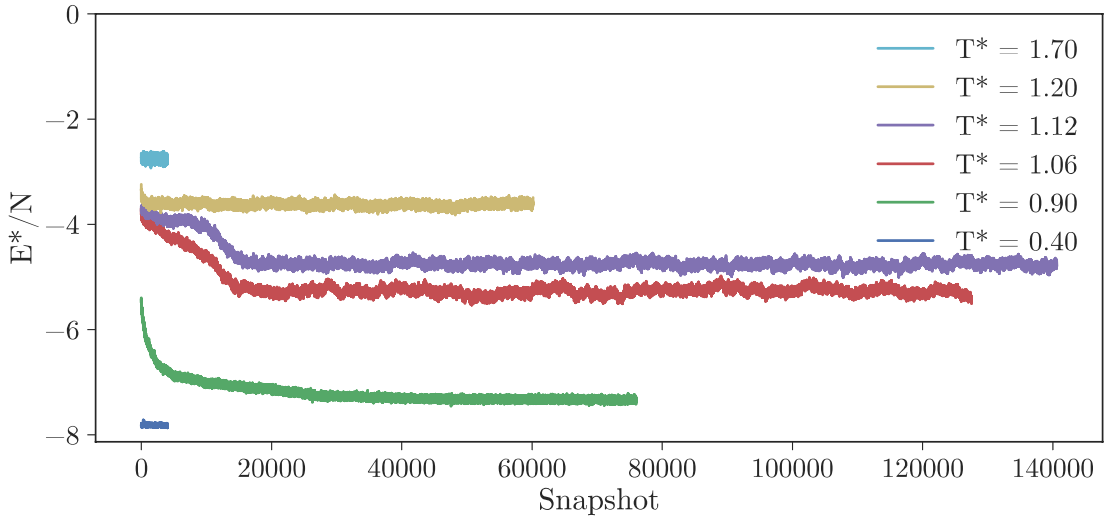


Figure 7.4: Evolution of reduced energy per particle for *trans*-azobenzene at six temperature values as simulation progresses. At $T^* = 1.16$, the simulation ran for the largest amount of sweeps, a total of 7030500. A snapshot was saved to disk every 50 sweeps, and the number of snapshots is used as the x-axis.

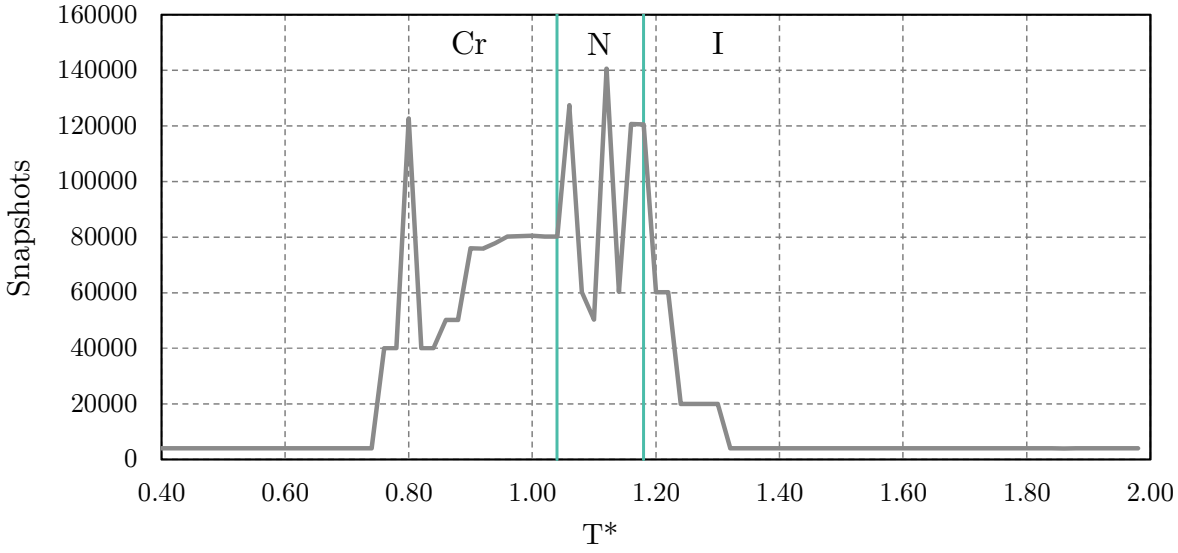


Figure 7.5: Number of snapshots in the *trans*-azobenzene simulations as a function of temperature. The two vertical lines denote phase transitions between crystal (Cr), nematic (N) and isotropic (I) phases, which are identified in Chapter 8.

Chapter 8

Results and discussion

8.1 Presenting the results

Trans-azobenzene production simulations were carried out at $P^* = 4.0$ and at a temperature range of $T^* = 0.40 \ldots 2.00$, $\Delta T^* = 0.02$. As stated in Section 7.3, full-scale *cis*-azobenzene production simulations were not carried out, so initial *cis* test simulation results are discussed in their place. Presented quantities are averages of individual values acquired from multiple consecutive snapshots. The values of thermodynamic quantities and the orientational order parameter P_2 were calculated as an average of thousands of snapshots (ranging from 2000 to around 20000 snapshots), depending on how fast the simulation reached equilibrium at the corresponding temperature. The translational order parameter τ_1 and the radial distribution functions were calculated using the 1000 final snapshots of the simulation.

Blocking method (see Section 5.8.2) was used to deduce whether the simulations had enough data, with a criterion of variance not changing in three consecutive blocking operations. Blocking method was also used to determine error bars for thermodynamic quantities and order parameters. Quantities such as energy, volume, pressure, temperature and distance are presented in reduced units (see Section 4.2.4).

In addition to single-site GB simulation studies [27, 62, 68], two two-site GB Monte Carlo studies of bent-core particles from 2002 act as reference studies here. Both articles employ an NPT ensemble, using parametrisations (3, 5, 1, 2) at $P^* = 3.0$ [51] and (3, 5, 2, 1) at $P^* = 2.0$ [52]. In the former article, an interesting Sm–N–I phase structure was discovered for a bending angle of 140°, and the latter article identified a multitude of exotic LC phases for different bending angles. The study also contains results for simulated rodlike two-site GB particles, which are compared to *trans*-azobenzene simulation results of this study.

8.2 *Trans*-azobenzene simulations

8.2.1 Order parameters

Orientational order parameter P_2 and translational order parameter τ_1 are presented in Fig. 8.1 as a function of simulation temperature. Error bars of both parameters are also displayed — their values are considerably small, especially when compared to amplitudes of irregularities in the parameter graphs. This could indicate that in some cases there is false precision present in the error calculated with the blocking method.

As the test simulations suggested, there are two clear phase transitions visible as discontinuities in P_2, near temperatures $T^* = 1.06$ and $T^* = 1.18$. At $T^* = 1.06 \ldots 1.18$ the value of the orientational order parameter $0.60 < P_2 < 0.87$ suggests a presence of a nematic phase there. At other temperatures, the value of P_2 stays almost constant: at $T^* \leq 1.04$, $P_2 > 0.97$ and at $T^* \geq 1.20$, $P_2 < 0.08$, which suggests an isotropic phase. In a reference study [52], a rodlike two-site GB model at $P^* = 2.0$ with parametrisation (3, 5, 2, 1) resulted in an orientational order parameter curve with a similar structure: the value quickly descends from a plateau of 0.951 to 0.115 as temperature rises from $T^* = 2.3$ to $T^* = 2.6$. The transition happens at a drastically different temperature, but this is expected as the parametrisation differs considerably.

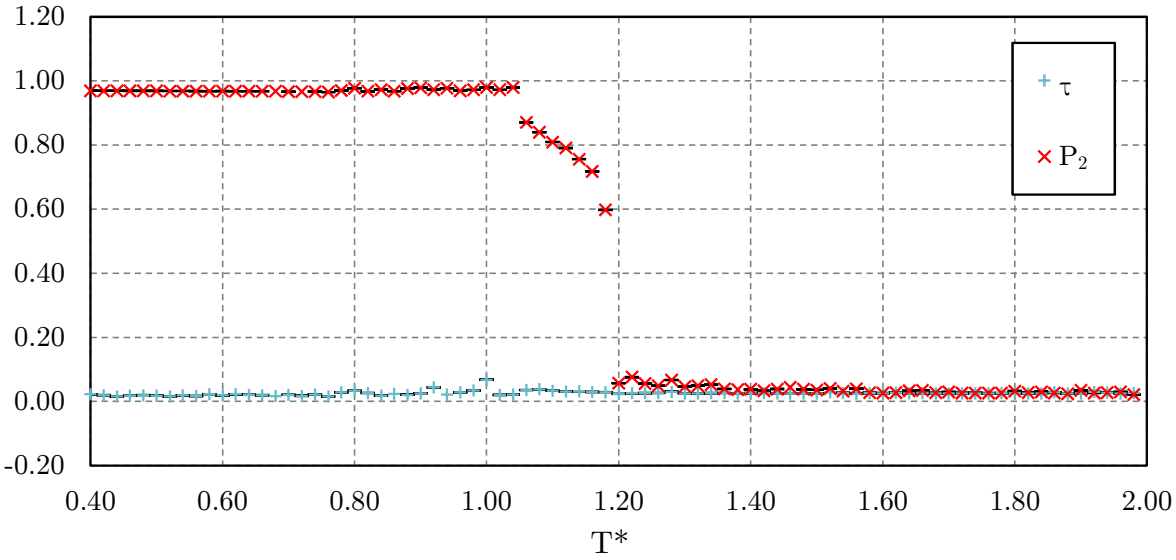


Figure 8.1: *Trans*-azobenzene orientational order parameter P_2 and translational order parameter τ_1 with error bars calculated with the blocking method.

The value of τ_1 stays close to zero at all temperatures. This suggests that a smectic-A phase is not present in these simulations, which is a noticeable difference to single-site GB simulations with the parametrisation (4.4, 20, 1, 1) by Karjalainen *et al.* [27, 30], and earlier single-site GB studies [62, 68]. Smectic phases have been identified even by using a rodlike two-site GB model [52] — this is discussed further in Section 8.2.3.

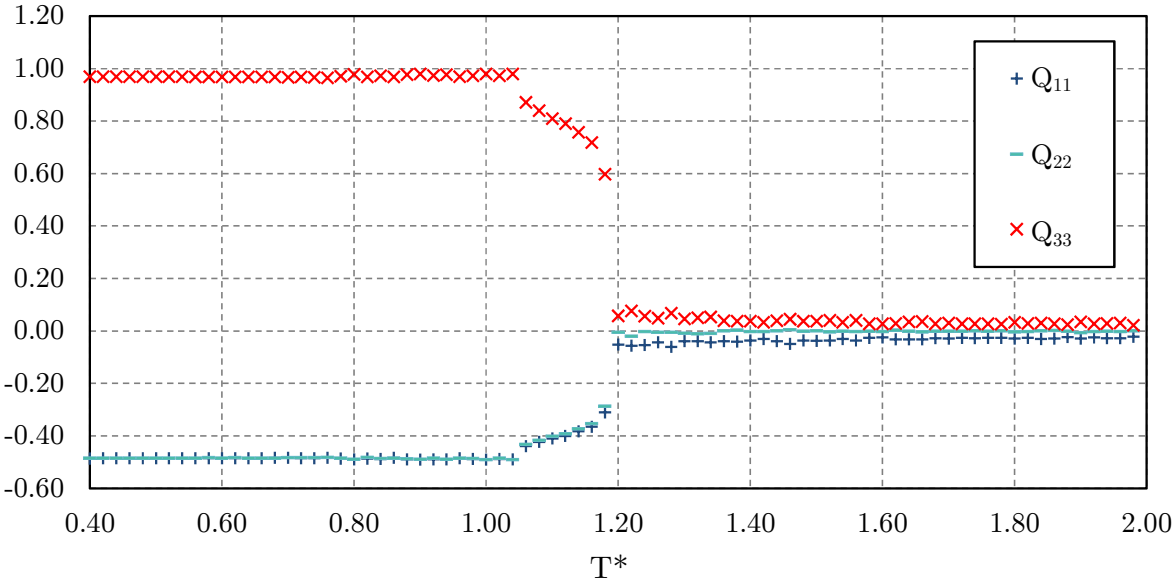


Figure 8.2: *Trans*-azobenzene eigenvalues for the orientational order parameter tensor Q, where Q_{33} is the orientational order parameter P_2.

Curiously, in the study by Karjalainen *et al.*, the translational order parameter had rather large values of $0.07 < \tau_1 < 0.15$ for a nematic phase. In this respect, the results in this study are more in line with earlier studies [62, 68], where $\tau_1 < 0.1$ in the bulk nematic phase.

All three eigenvalues of the orientational tensor $\boldsymbol{Q}$ (Eq. (3.1)) are presented in Fig. 8.2. Interestingly, even in the isotropic phase, there is a visible difference between the tensor elements — consistently, $Q_{11} \approx -Q_{33}$. Measure of biaxiality T was calculated with Eq. (3.4) and the results are seen in Fig. 8.3. A hint of biaxiality is visible at temperatures $T^* > 1.0$. This is unexpected, as the axially symmetric model should not result in biaxial phases — thus, this might just be a result of numerical noise or fluctuations due to a finite system size.

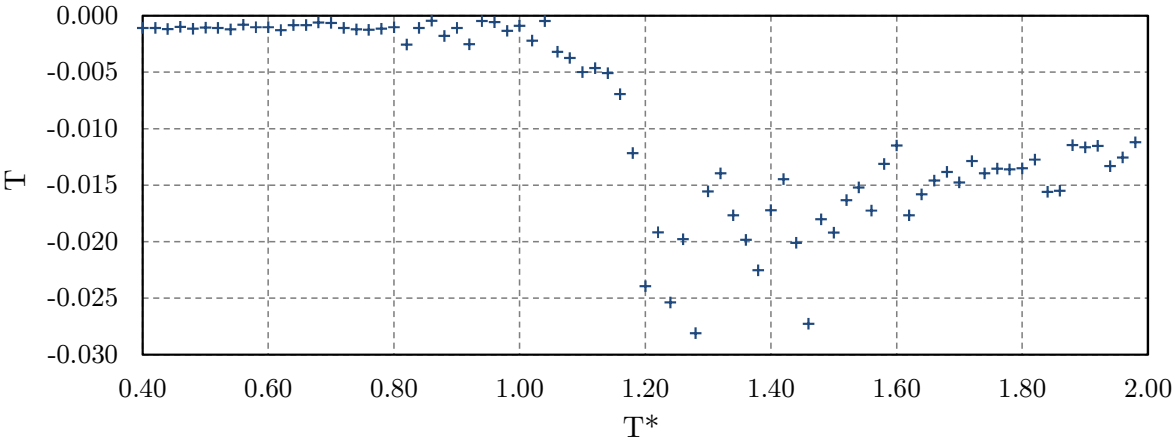


Figure 8.3: *Trans*-azobenzene measure of biaxiality T.

8.2.2 Thermodynamic quantities

Reduced energy and enthalpy per particle are presented in Figures 8.4 and 8.5. Number density, or the number of particles per volume, is shown in Fig. 8.6. Akin to order parameter P_2, the phase transitions are visible as discontinuities in all three graphs: at $T^* = 1.06 \rightarrow 1.04$, there is a large jump in energy, enthalpy and density. When comparing to values at $T^* = 1.06$, there is a relative density change of 8.53%, enthalpy change of -29.7% and energy change of -34.6%. The jump at $T^* = 1.20 \rightarrow 1.18$ is much less pronounced. Also, near the phase transitions, the magnitude of error in enthalpy and energy is especially large: e.g., at $T^* = 1.18$, $\Delta E^*/E^* = 33\%$. In general, the magnitudes of error are at their largest in the nematic phase, where simulations generally took longer to equilibrate. In contrast, though, number density error is diminutive at all temperatures. Like the orientational order parameter in Section 8.2.1, the number density graph resembles its counterpart in a previous two-site GB study [52]. In this study, however, the discontinuity at $T^* = 1.06$ is much more pronounced than in the reference study, where the transition phase between a highly ordered phase and isotropic was more linear.

Additionally, there are slight irregularities in the thermodynamic quantity values at temperatures $T^* = 0.78 \ldots 1.04$, which suggests the presence of some unidentified phase structure there. This is supported by a long equilibration period at $T^* = 0.80$, as seen in Fig. 7.5. The irregularities are highlighted in Fig. 8.6. Oppositely, at temperatures $T^* < 0.80$, the temperature dependence of thermodynamic quantities was almost completely linear. The irregularities and the linear region are also faintly visible in the orientational order parameter P_2 in Fig. 8.1.

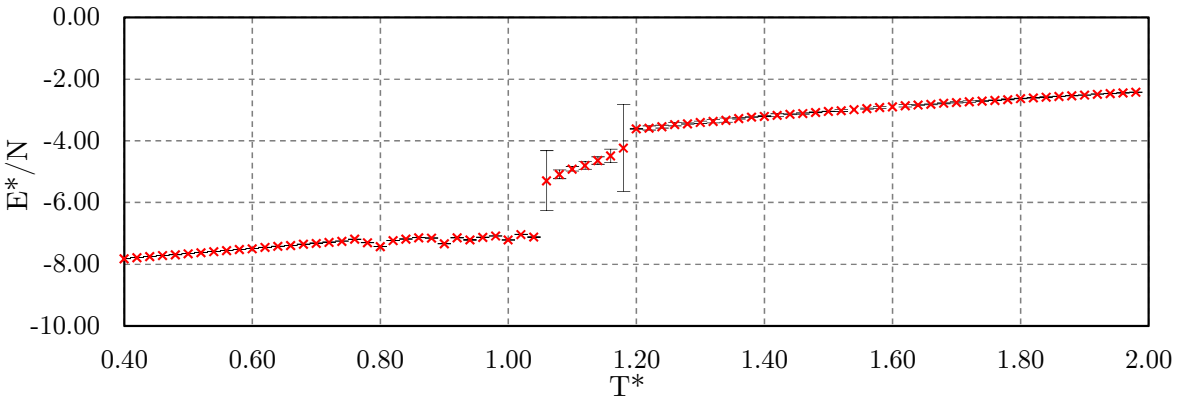


Figure 8.4: *Trans*-azobenzene reduced energy per particle at $P^* = 4.0$ with error bars.

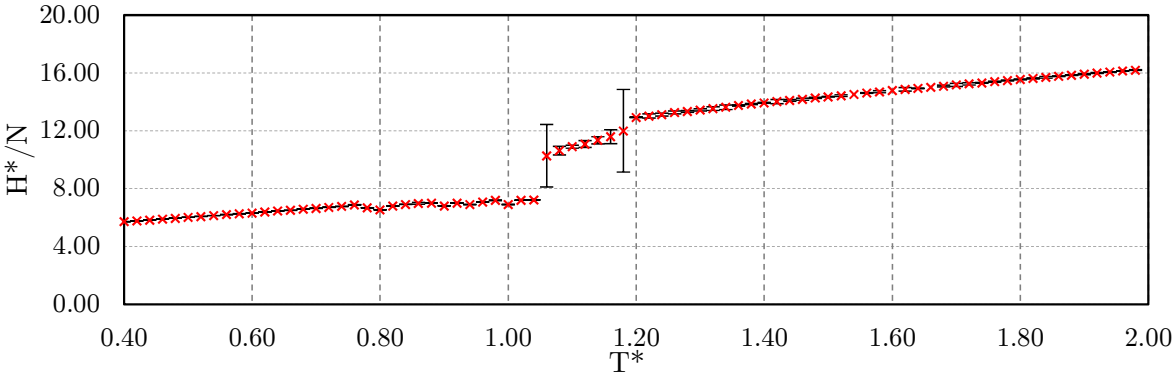


Figure 8.5: *Trans*-azobenzene reduced enthalpy per particle at $P^* = 4.0$ with error bars.

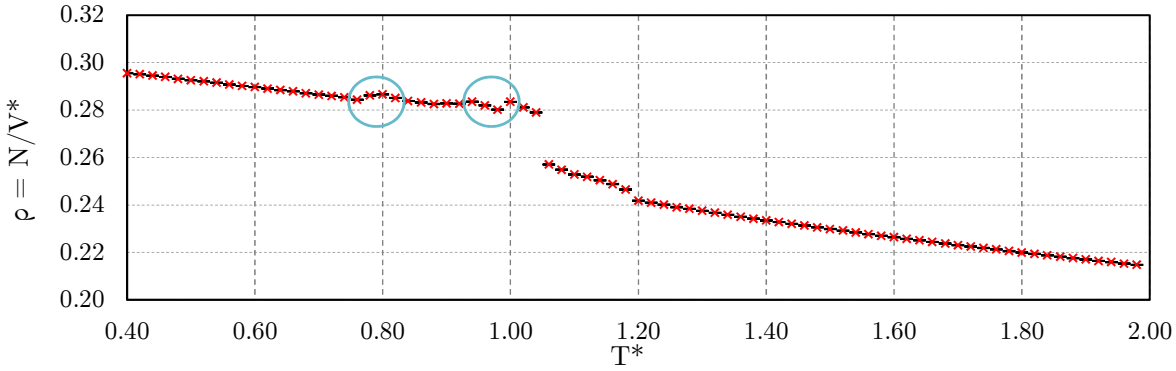


Figure 8.6: *Trans*-azobenzene number density at $P^* = 4.0$ with error bars. Irregularities at $T^* = 0.80$ and $T^* = 0.98$ are highlighted.

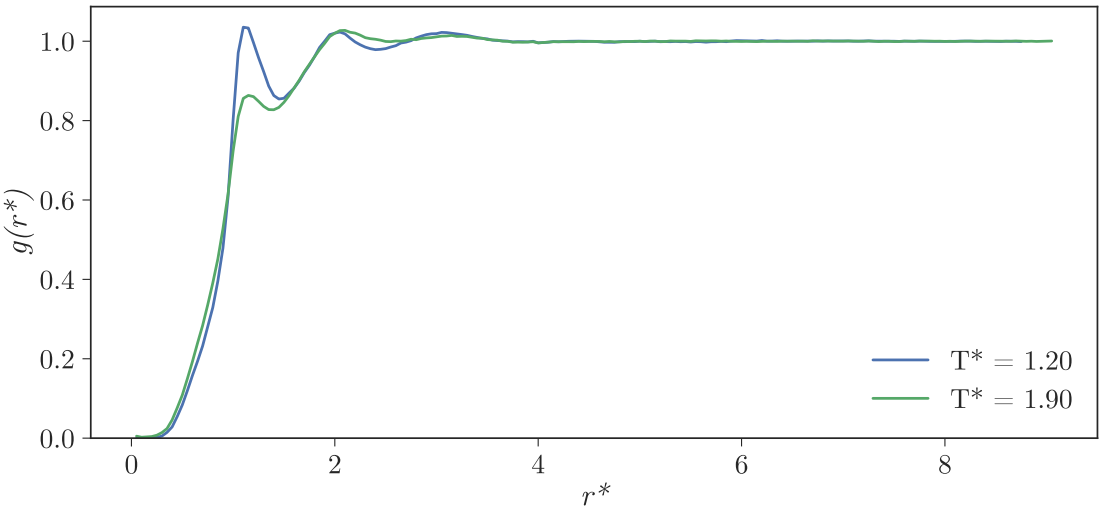
8.2.3 Radial distribution functions

Radial distribution functions (RDFs) were calculated with an accuracy of $\Delta r^* = 0.05$. At high temperatures, as seen in Fig. 8.7a, there are only three small peaks visible in RDFs, where the distance to the nearest neighbour is $r^* = 1.15$ at $T^* = 1.90$ and $r^* = 1.10$ at $T^* = 1.20$. At distances $r^* > 4$, correlations disappear completely and the graphs steadily decay to a constant value of one. The three-peak RDF shape suggests the presence of a liquid phase (see Fig. 3.2 for comparison). In the nematic phase (Fig. 8.7b), the distance to the nearest neighbour slightly increases as temperature drops: $r^* = 1.20$ at $T^* = 1.06$. Additionally, the peaks designating the next neighbours become more pronounced at lower temperatures, and long-range order starts to persist.

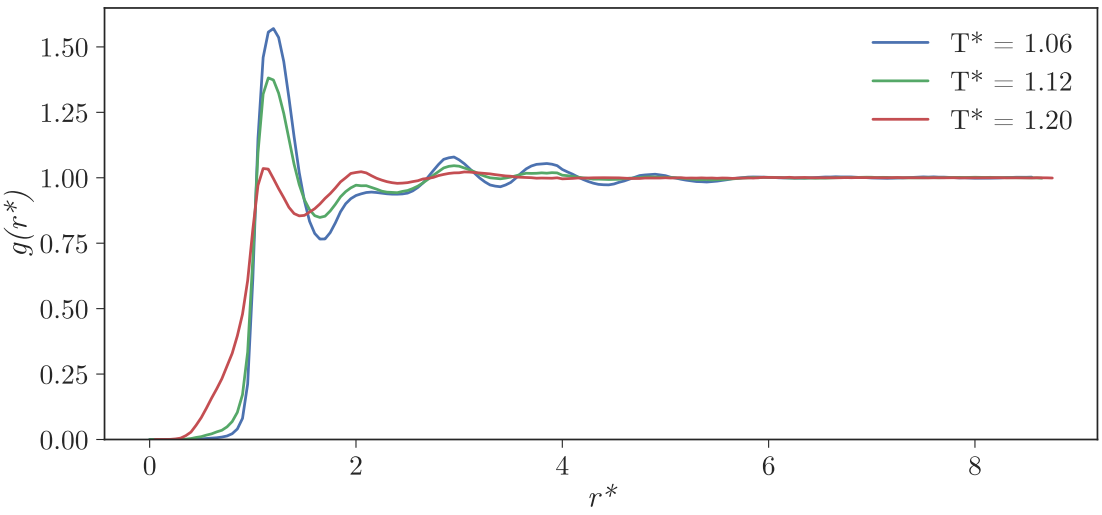
At the $T^* = 1.06 \rightarrow 1.04$ transition, there is a significant transformation in the RDF. As seen in Fig. 8.7c, the nearest neighbour peak moves further away and grows taller and sharper at $T^* = 1.04$, $r^* = 1.35$. Simultaneously, a smaller peak appears at $r^* = 1.10$. The peaks of the next neighbours become even more pronounced. This suggests a crystal phase at temperatures $T^* \leq 1.04$, as sharp peaks usually characterise crystal phases [54]. Fig. 8.7d shows that at lower temperatures, the sharp features become even sharper, more peaks appear and the gap between the highest peak and

the closer sub-peak deepens. This supports the argument for the presence of a crystal phase.

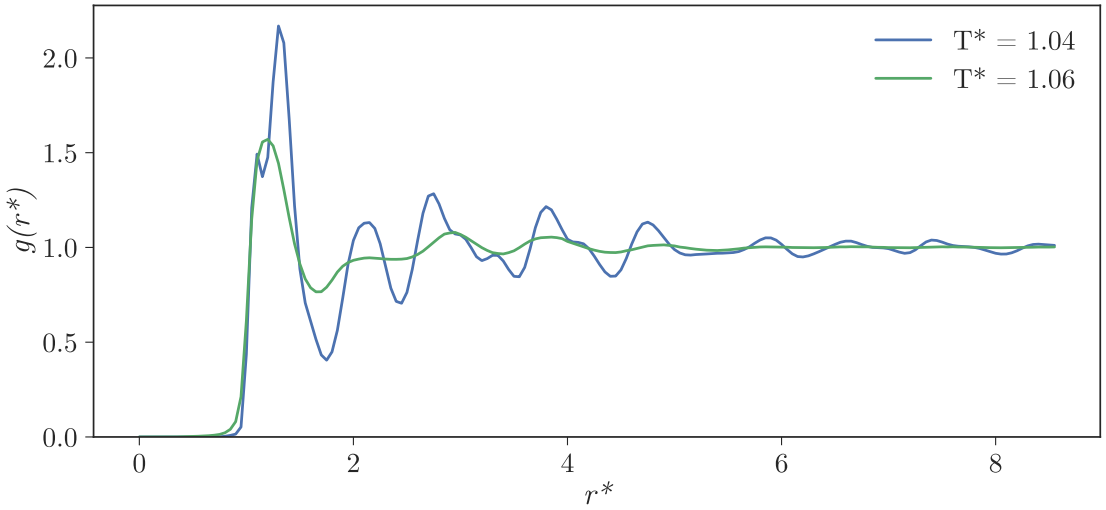
In a reference study [52], in addition to nematic and isotropic phases, smectic-A and smectic-B phases were identified for a rodlike two-site GB model at $P^* = 2.0$, $T^* < 2.3$ with parametrisation (3, 5, 2, 1). The phase was indicated by a split of the second peak in the pair distribution function. Such splits are not visible at any temperatures here. This supports the translational order parameter results discussed in Section 8.2.1, suggesting that a smectic phase is not present here. This could be caused by dissimilar parametrisations, or a difference in the site distance: having slightly overlapping GB particles makes the reference study model resemble the single-site GB model to a greater extent than the dimers with a more pronounced gap in this study do.



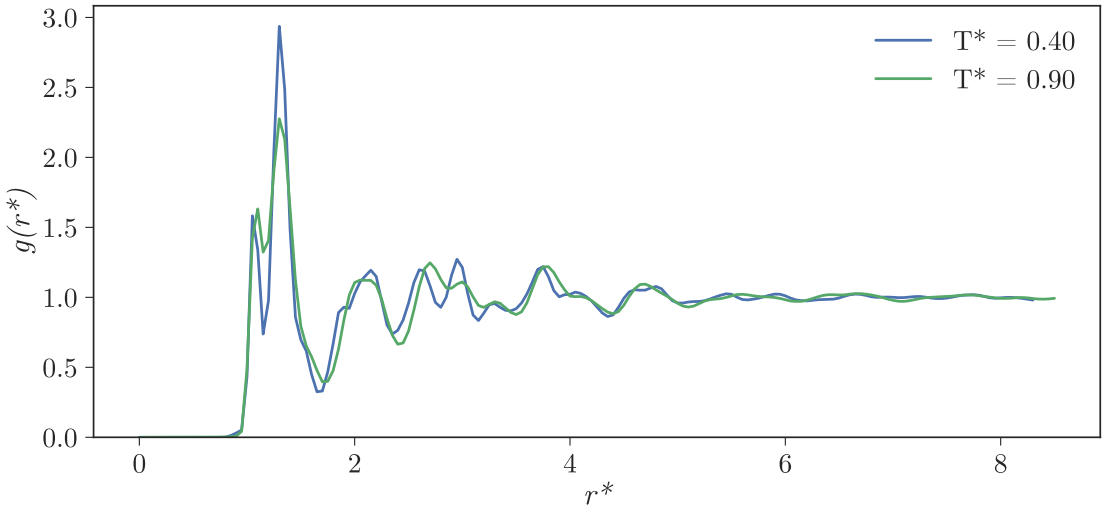
(a) *Trans*-azobenzene radial distribution functions in the isotropic phase.



(b) *Trans*-azobenzene radial distribution functions in the nematic phase.



(c) *Trans*-azobenzene radial distribution functions in the phase transition $T^* = 1.06 \rightarrow 1.04$.



(d) *Trans*-azobenzene radial distribution functions in the crystalline phase.

Figure 8.7: *Trans*-azobenzene radial distribution functions.

8.3 *Cis*-azobenzene simulations

Reduced enthalpy per particle from the initial *cis*-azobenzene test simulations was already presented in the previous chapter in Fig. 7.3. Here, reduced energy per particle and number density are shown in Figures 8.8 and 8.9, respectively. The temperature dependence of energy, enthalpy and density is monotonic at all three simulated pressure values, and thus, there seems to be no interesting phase transitions present.

The vector used to calculate the orientational order parameter P_2 was the difference vector of the GB site orientation vectors $\hat{\boldsymbol{u}} - \hat{\boldsymbol{v}}$, as stated in Section 3.3.1. The values of P_2 at three simulated pressures is presented in Fig. 8.10. The value of P_2 was less than 0.04 at all temperature points, and at $P^* = 60$ less than 0.02. This confirms that a nematic phase does not appear in the *cis* simulation. The translational order parameter was also calculated, but unsurprisingly its value was $\tau_1 < 0.03$ at all temperatures, so a smectic phase is not present, either.

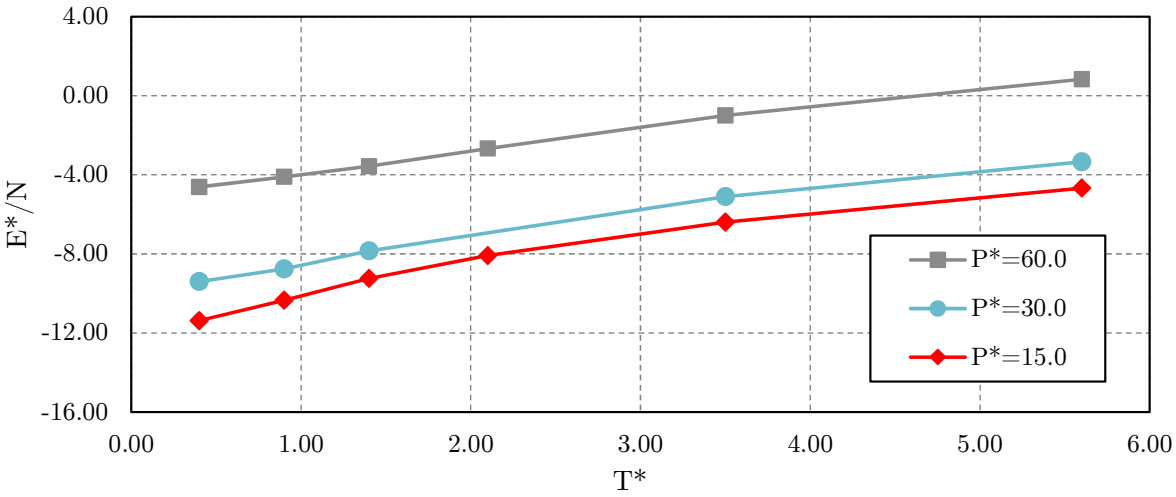


Figure 8.8: *Cis*-azobenzene reduced energy per particle.

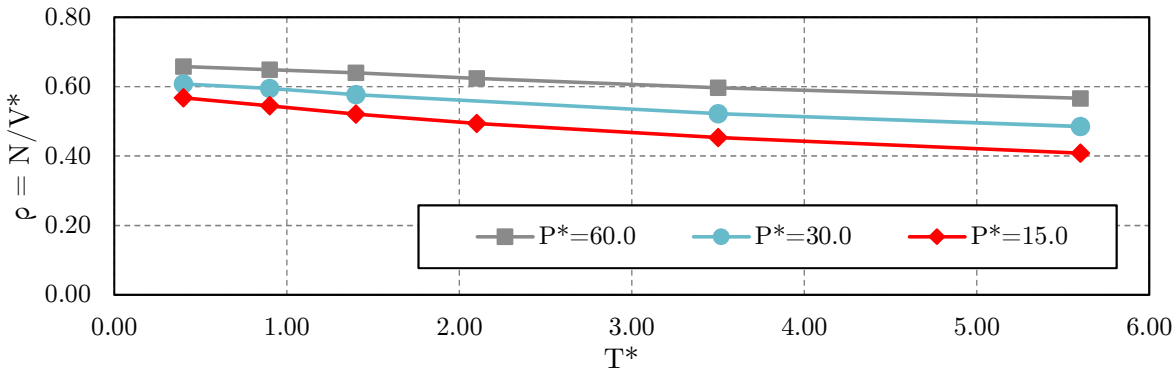


Figure 8.9: *Cis*-azobenzene number density in test simulations.

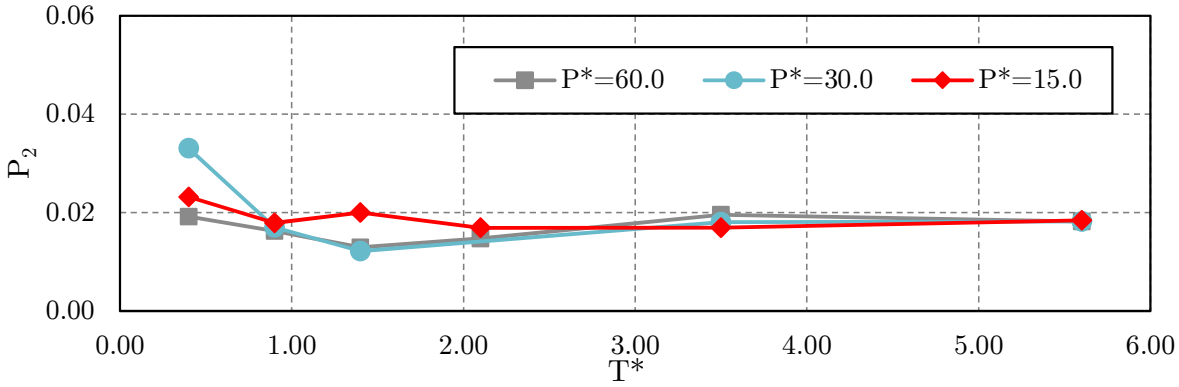


Figure 8.10: *Cis*-azobenzene test simulation orientational order parameter P_2 in test simulations.

Radial distribution functions at the minimum and maximum simulated temperatures are presented in Fig. 8.11. At $T^* = 5.6$ there are almost no distinguishing features at all. This is unusual; even in isotropic phases, there should be at least one peak present in the RDF. The lack of peaks could result from the hard core of the potential discussed in Section 7.2. At lower temperatures, namely at $T^* = 0.4$, some peaks have formed, the highest one being at $r^* = 1.3$. At $P^* = 60$, there is an additional nearest neighbour peak at $r^* = 0.6$. Sharpness of the peaks suggests the presence of a crystalline phase, which would in turn suggest that there is a phase transition from liquid to solid somewhere between the minimum and maximum temperatures. As there are no stark changes in the graphs, the exact temperature is difficult to pinpoint.

Previous two-site GB studies have found more interesting phase diagrams for bent-core particles: a parametrisation (3, 5, 1, 2) has produced a Sm–N–I phase structure for a GB dimer with a bending angle of 140° [51], the untilted smectic phase even containing global phase biaxiality. Another study with a parametrisation (3, 5, 2, 1) simulated dimer models with bending angles between 110° and 180° [52], finding that the phase structure is starkly influenced by the bending angle: a rodlike (180°) model formed isotropic, nematic, smectic-A and smectic-B phases. In the bent-core models with bending angles $120° - 180°$, at least one ordered phase was detected, varying between SmA, SmB, N and a TGB-like ("twist-grain boundary") phase at different bending angle values. For 110°, no ordered phases were found.

The lack of ordered phases for *cis*-azobenzene in this study can be attributed to two factors: in the *cis* parametrisation, the GB length-to-breadth ratio κ had a rather small value of ∼1.22, which made the particles almost spherical. On the other hand, the sharp bending angle of ∼71° made the length-to-breadth ratio of the entire dimer ∼1.55, which is quite close to one as well. In the two-site study mentioned above, it was argued that for particles with $\kappa = 3$, even a bending angle of 110° was deemed too small for liquid crystal phases, so it is no wonder that no LC phases were found with the parametrisation used in this study, either.

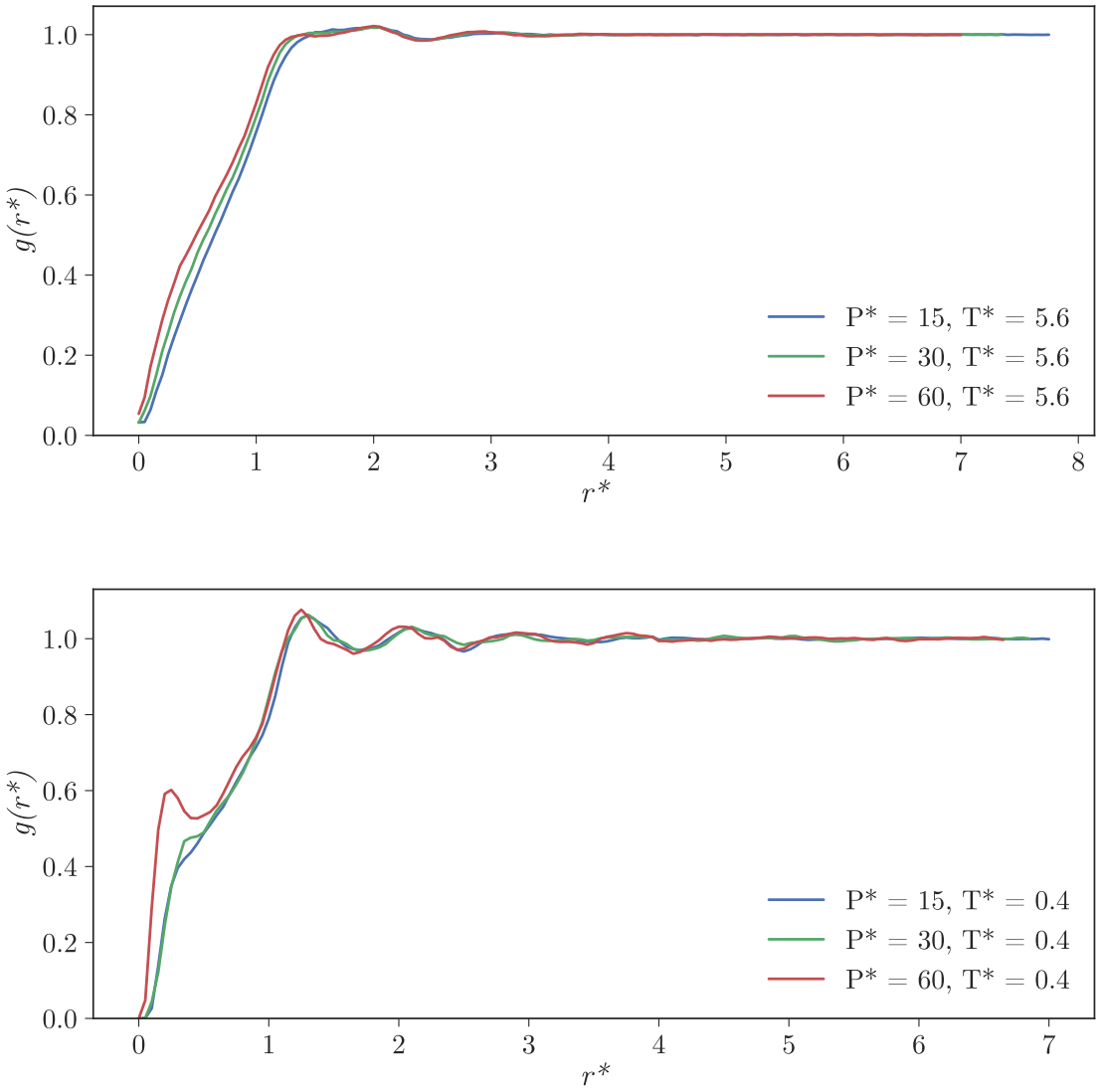


Figure 8.11: *Cis*-azobenzene radial distribution functions at temperatures $T^* = 5.6$ and $T^* = 0.4$.

8.4 Visualisations

Simulation snapshot visualisations were created with Blender 3D and a dedicated Python script (the same method was earlier used for visualising initial states in Fig. 7.1). Final snapshots of several simulations were visualised, out of which six TAB snapshots around the nematic phase temperature and four CAB snapshots are presented in Figures 8.13 and 8.12, respectively.

The CAB snapshots at $P^* = 60$ look isotropic and chaotic, even. The only tangible difference between different temperatures is the simulation box size: it is smaller at cooler temperatures, which means density is higher. At $T^* = 0.4$, $\rho = 0.658$ and at $T^* = 5.6$, $\rho = 0.566$.

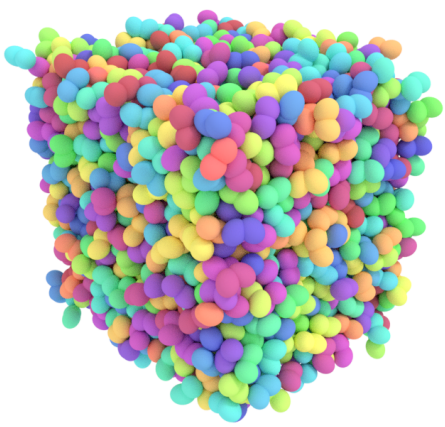
By contrast, in the TAB snapshots the phase transition is clearly visible: at lower temperatures, particles are more aligned towards a common axis. This axis — the director — is not exactly aligned to the global z-axis, but it is significantly more aligned to z-axis than in some initial simulations where anisotropic scaling was not applied. E.g., at temperature $T^* = 0.80$, the director average of the pictured final snapshot of TAB is $-0.186\hat{\boldsymbol{i}} - 0.246\hat{\boldsymbol{j}} + 0.942\hat{\boldsymbol{k}}$. To illustrate the density difference around the nematic phase, at $T^* = 0.80$, $\rho = 0.287$ and at $T^* = 1.30$, $\rho = 0.237$.

8.5 Phase structure

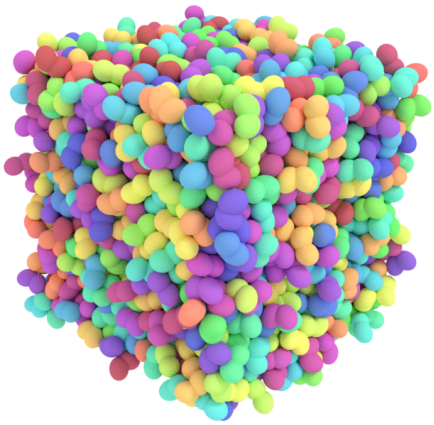
Based on the presented data, the phase structure is most likely Cr-N–I for *trans*-azobenzene:

$$\mathrm{Cr} \underset{1.04}{\longleftarrow} \mathrm{N} \underset{1.18}{\longleftarrow} \mathrm{I} \tag{8.1}$$

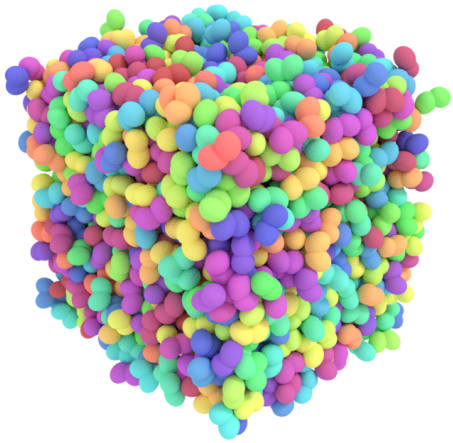
Here, the values under the arrows are the transition temperatures in the cooling series in reduced units. There was also a possible unidentified phase transition at $T^* = 0.76$, signified by a slight transition in thermodynamic quantities and the orientational order parameter. The simulated model differs from earlier similar single-site studies [27, 62, 68] by lacking the SmA phase, and from a previous two-site study [52] by lacking the SmB and SmA phases. For *cis*-azobenzene, the phase structure is presumably Cr–I, with an indeterminate transition temperature.



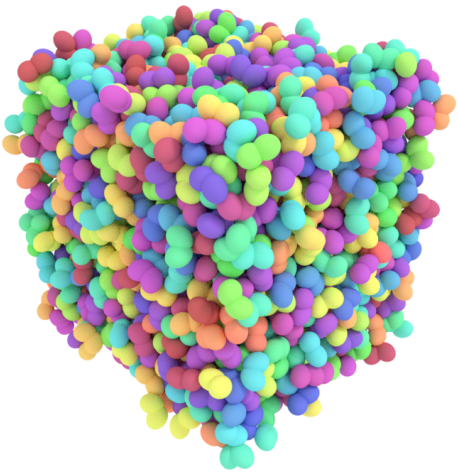
(a) T*=0.40



(b) T*=1.40

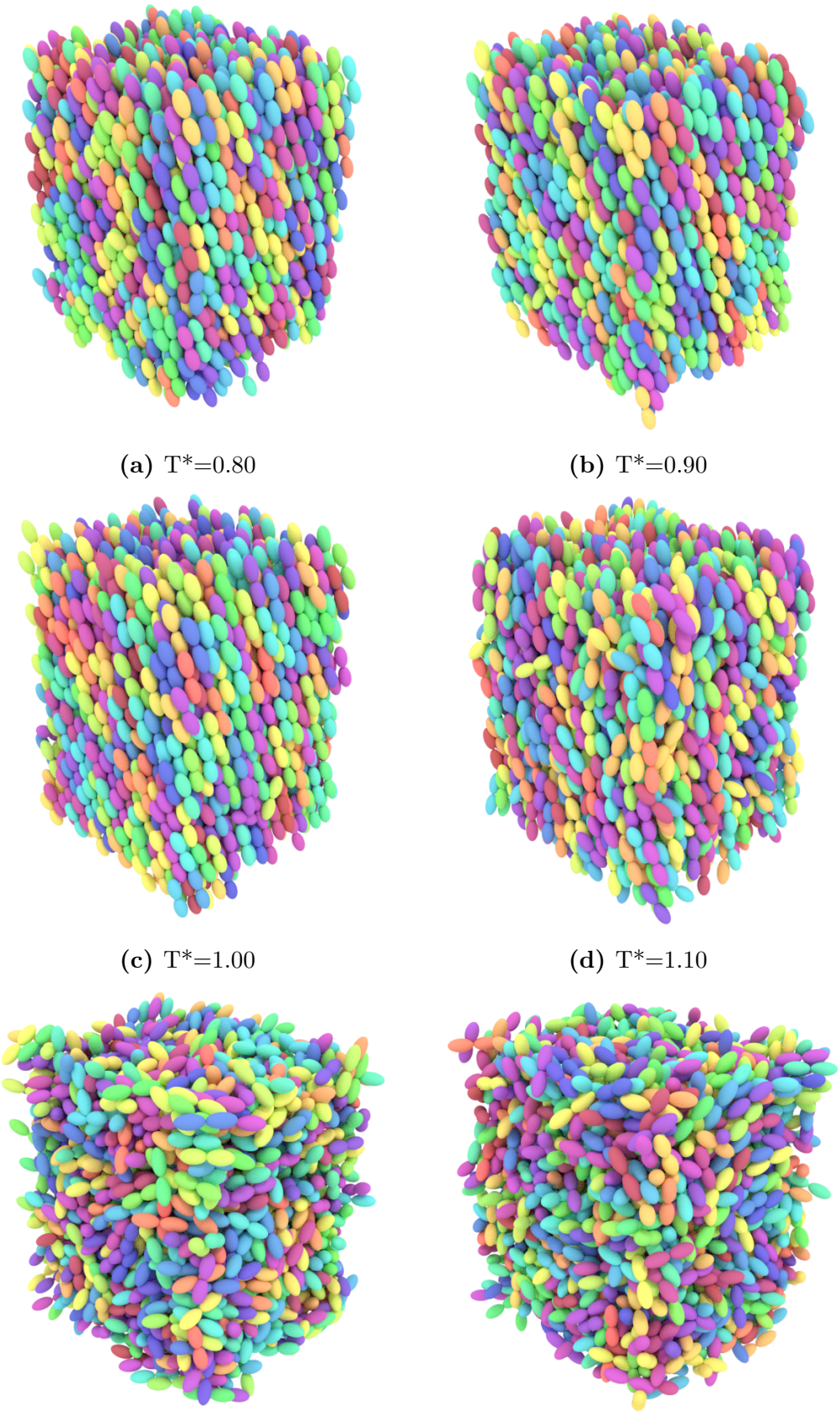


(c) T*=3.50



(d) T*=5.60

Figure 8.12: Final snapshots of *cis*-azobenzene test simulations at pressure $P^* = 60$.



(a) T*=0.80 **(b)** T*=0.90

(c) T*=1.00 **(d)** T*=1.10

(e) T*=1.20 **(f)** T*=1.30

Figure 8.13: Final snapshots of *trans*-azobenzene simulations at six temperature values, pressure $P^* = 4.0$.

Chapter 9

Conclusions

In this thesis, I built two two-site Gay–Berne model parametrisations (1.22, 2.01, 1, 1) and (1.92, 18.63, 1, 1) to depict *cis*- and *trans*-azobenzene molecules, respectively. The parametrisations were created by fitting a rigid two-site GB model to quantum-chemical azobenzene data with the least squares method. Quantum-chemical interaction energies were carried out at the SCS-MP2 and def2-TZVPP basis set level. The particles were simulated in bulk ($N = 2000$) by means of isothermal-isobaric Monte Carlo in the PTGBCYL simulation software [27]. Full-scale production simulations were only carried out for the *trans* molecule, as the *cis* parametrisation seemed to create a somewhat featureless phase structure, based on results from the initial test simulations. Thermodynamic quantities and structural liquid crystal parameters were calculated as a function of temperature, and the results were compared to earlier single-site Gay–Berne simulations [27, 30] with a parametrisation (4.4, 20, 1, 1) and two-site Gay–Berne studies with parametrisations (3, 5, 1, 2) [51] and (3, 5, 2, 1) [52].

Phase structure was determined with the calculated quantities, and the results revealed a Cr–N–I phase structure with well-defined transition temperatures for *trans*-azobenzene, and a Cr–I structure with no clear transition temperature for *cis*-azobenzene. *Trans* simulation results were somewhat expected based on previous one-site GB studies with a similar parametrisation [27], but a significant divergence here was the absence of any smectic phases, indicated by a comparatively small translational order parameter value. One possible explanation for this divergence could be the rather short and cumbersomely applied cutoff radius of the potential discussed in Section 7.2. In previous studies, smectic-A and smectic-B phases have been identified by using a similar one-site GB parametrisation [62, 68] and even by using a rodlike two-site GB model [52].

The *cis* simulation results turned out lacklustre for two reasons. Firstly, the molecule itself is relatively short and has a sharper bending angle of $\sim 71^\circ$ than many LC-forming bent-core molecules. LC-forming bent-core molecules with azo and benzene groups [72, 73] and bent-core molecules in general [74, 75] have been studied exten-

sively, and the bending angle is usually 100° or greater. Secondly, the least squares fit resulted in a parametrisation with a length-to-breadth ratio of ∼1.22, making the GB particles almost spherical. These two factors made the model very unlikely to form any LC phases in the first place. Then again, experimental results [1, 35] have discovered LC phases only for the *trans* isomer of azobenzene.

One missed target of this study was creating a single parametrisation that would have depicted both *cis* and *trans* conformations by using a different bending angle. As a consequence, mixed *cis*-*trans* simulations became arduous to implement in PTGBCYL, and only all-*cis* and all-*trans* simulations were carried out. As another slight remark, duration of the production runs was relatively long as parallelisation was not utilised in this study. Long durations could be countered by using a larger step size, e.g., $\Delta T^* = 0.25$, to reduce the number of simulated temperature points. This could, however, obscure some of the finer details in the phase diagram.

Possible continuation studies for the two-site model include simulating azobenzene derivatives with a larger breadth-to-width ratio, in order to see if longer *cis* forms support a more interesting phase structure. Simulating mixed *cis*–*trans* ensembles with a modified software or a single parametrisation would open up new possibilities. Even flexible bent-core models have been simulated before [76] — these might be applicable for simulating transitions between the conformations, e.g., for studying the thermal relaxation from *cis*- to *trans*-azobenzene.

Bibliography

[1] T. Ikeda & O. Tsutsumi: Optical switching and image storage by means of azobenzene liquid-crystal films, *Science*, **268**:5219 (1995), 1873–1875.

[2] J. Gay & B. Berne: Modification of the overlap potential to mimic a linear site-site potential, *The Journal of Chemical Physics*, **74**:6 (1981), 3316–3319.

[3] P. Atkins & R. Friedman: *Molecular Quantum Mechanics*, Fifth Edition, Oxford University Press, 2011.

[4] D. R. Hartree: The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods, *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**:1 (1928), 89–110.

[5] V. Fock: Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, *Zeitschrift für Physik*, **61**:1 (1930), 126–148.

[6] D. K. W. Mok, R. Neumann, & N. C. Handy: Dynamical and nondynamical correlation, *The Journal of Physical Chemistry*, **100**:15 (1996), 6225–6230.

[7] C. Møller & M. S. Plesset: Note on an approximation treatment for many-electron systems, *Physical Review*, **46**:7 (1934), 618–622.

[8] J. Čížek: On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods, *The Journal of Chemical Physics*, **45**:11 (1966), 4256–4266.

[9] P. Peuravaara: *Temperature-dependent chemical shift in the aqueous solution of xenon*, Master's thesis, University of Oulu, 2017.

[10] P. Lantto: *NMR-parametrien laskeminen ab initio -menetelmillä*, Master's thesis, University of Oulu, 1998.

[11] O. Christiansen, H. Koch, & P. Jørgensen: The second-order approximate coupled cluster singles and doubles model CC2, *Chemical Physics Letters*, **243**:5 (1995), 409–418.

[12] S. Grimme: Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies, *The Journal of Chemical Physics*, **118**:20 (2003), 9095–9102.

[13] Y. Jung, R. C. Lochan, A. D. Dutoi, & M. Head-Gordon: Scaled opposite-spin second order Møller–Plesset correlation energy: An economical electronic structure method, *The Journal of Chemical Physics*, **121**:20 (2004), 9793–9802.

[14] *Turbomole 6.5 User's Manual*, COSMOlogic, 2013, URL: `http://www.cosmologic.de/data/chemistry/DOC.pdf` (visited on 08/19/2013).

[15] J. C. Slater: Atomic shielding constants, *Physical Review*, **36**:1 (1930), 57–64.

[16] A. Schäfer, H. Horn, & R. Ahlrichs: Fully optimized contracted Gaussian basis sets for atoms Li to Kr, *The Journal of Chemical Physics*, **97**:4 (1992), 2571–2577.

[17] T. Hansson: *Theoretical study of the properties of As_nSb_m, As_nBi_m and Sb_nBi_m $(n+m=2,4)$ binary clusters*, Master's thesis, University of Oulu, 2015.

[18] *Q-Chem 4.3 User's manual: Auxiliary basis set (resolution-of-identity) MP2 methods*, Q-Chem, Inc., 2018, URL: `http://www.q-chem.com/qchem-website/manual/qchem43_manual/sect-auxRI.html` (visited on 03/04/2019).

[19] M. Feyereisen, G. Fitzgerald, & A. Komornicki: Use of approximate integrals in ab initio theory. An application in MP2 energy calculations, *Chemical Physics Letters*, **208**:5 (1993), 359–363.

[20] T. H. Dunning Jr.: Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *The Journal of Chemical Physics*, **90**:2 (1989), 1007–1023.

[21] R. Ditchfield, W. J. Hehre, & J. A. Pople: Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules, *The Journal of Chemical Physics*, **54**:2 (1971), 724–728.

[22] P. Lantto & J. Vaara: *761668S Computational physics and chemistry, lecture overview*, University of Oulu, 2016.

[23] S. Boys & F. Bernardi: The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors, *Molecular Physics*, **19**:4 (1970), 553–566.

[24] P. J. Collings & M. Hird: *Introduction to Liquid Crystals: Chemistry and Physics*, Taylor & Francis Ltd., 1997.

[25] A. Cuetos, A. Galindo, & G. Jackson: Thermotropic biaxial liquid crystalline phases in a mixture of attractive uniaxial rod and disk particles, *Physical Review Letters*, **101**:23 (2008), 237802.

[26] J. Vieillard-Baron: The equation of state of a system of hard spherocylinders, *Molecular Physics*, **28**:3 (1974), 809–818.

[27] J. Karjalainen, J. Lintuvuori, V.-V. Telkki, P. Lantto, & J. Vaara: Constant-pressure simulations of Gay–Berne liquid-crystalline phases in cylindrical nanocavities, *Physical Chemistry Chemical Physics*, **15**: (2013), 14047.

[28] R. J. Low: Measuring order and biaxiality, *European Journal of Physics*, **23**:2 (2002), 111–117.

[29] M. Bates & G. Luckhurst: Computer simulation studies of anisotropic systems. XXX. The phase behavior and structure of a Gay–Berne mesogen, *The Journal of Chemical Physics*, **110**:14 (1999), 7087–7108.

[30] J. Karjalainen: *Gay–Berne-nestekiteen rakenne ja järjestys sylinterinmuotoisessa nanokaviteetissa*, Master's thesis, University of Oulu, 2009.

[31] M. Allen & D. Tildesley: *Computer Simulation of Liquids*, Oxford University Press, 1987.

[32] D. Chandler: *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.

[33] M. Pederzoli, J. Pittner, M. Barbatti, & H. Lischka: Nonadiabatic molecular dynamics study of the cis–trans photoisomerization of azobenzene excited to the S1 state, *The Journal of Physical Chemistry A*, **115**:41 (2011), 11136–11143.

[34] J. M. Ilnytskyi, D. Neher, & M. Saphiannikova: Opposite photo-induced deformations in azobenzene-containing polymers with different molecular architecture: Molecular dynamics study, *The Journal of Chemical Physics*, **135**:4 (2011), 044901.

[35] O. Tsutsumi, T. Shiono, T. Ikeda, & G. Galli: Photochemical phase transition behavior of nematic liquid crystals with azobenzene moieties as both mesogens and photosensitive chromophores, *The Journal of Physical Chemistry B*, **101**:8 (1997), 1332–1337.

[36] D.-Y. Kim, S.-A. Lee, H. Kim, S. Min Kim, N. Kim, & K.-U. Jeong: An azobenzene-based photochromic liquid crystalline amphiphile for a remote-controllable light shutter, *Chemical Communications*, **51**:55 (2015), 11080–11083.

[37] A. Marini, B. Zupančič, V. Domenici, B. Mennucci, B. Zalar, & C. A. Veracini: A photosensitive liquid crystal studied by ^{14}N NMR, ^{2}H NMR, and DFT calculations, *ChemPhysChem*, **13**:17 (2012), 3958–3965.

[38] D. Sainova, A. Zen, H.-G. Nothofer, U. Asawapirom, U. Scherf, R. Hagen, T. Bieringer, S. Kostromine, & D. Neher: Photoaddressable alignment layers for fluorescent polymers in polarized electroluminescence devices, *Advanced Functional Materials*, **12**:1 (2002), 49–57.

[39] R. H. Berg, S. Hvilsted, & P. S. Ramanujam: Peptide oligomers for holographic data storage, *Nature*, **383**:6600 (1996), 505–508.

[40] H. S. Lim, J. T. Han, D. Kwak, M. Jin, & K. Cho: Photoreversibly switchable superhydrophobic surface with erasable and rewritable pattern, *Journal of the American Chemical Society*, **128**:45 (2006), 14458–14459.

[41] K. Ichimura, S.-K. Oh, & M. Nakagawa: Light-driven motion of liquids on a photoresponsive surface, *Science*, **288**:5471 (2000), 1624–1626.

[42] H. Finkelmann, E. Nishikawa, G. G. Pereira, & M. Warner: A new opto-mechanical effect in solids, *Physical Review Letters*, **87**:1 (2001), 015501.

[43] M. R. Wilson: Molecular simulation of liquid crystals: progress towards a better understanding of bulk structure and the prediction of material properties, *Chem. Soc. Rev.* **36**:12 (2007), 1881–1888.

[44] C. Peter & K. Kremer: Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back, *Soft Matter*, **5**:22 (2009), 4357–4366.

[45] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, & A. Kolinski: Coarse-grained protein models and their applications, *Chemical Reviews*, **116**:14 (2016), 7898–7936.

[46] J. E. Jones & S. Chapman: On the determination of molecular fields. II. From the equation of state of a gas, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **106**:738 (1924), 463–477.

[47] B. J. Berne & P. Pechukas: Gaussian model potentials for molecular interactions, *The Journal of Chemical Physics*, **56**:8 (1972), 4213–4216.

[48] G. R. Luckhurst, R. A. Stephens, & R. W. Phippen: Computer simulation studies of anisotropic systems. XIX. Mesophases formed by the Gay-Berne model mesogen, *Liquid Crystals*, **8**:4 (1990), 451–464.

[49] E. D. Miguel, L. F. Rull, M. K. Chalam, & K. E. Gubbins: Liquid crystal phase diagram of the Gay-Berne fluid, *Molecular Physics*, **74**:2 (1991), 405–424.

[50] J. T. Brown, M. P. Allen, E. Martín del Río, & E. d. Miguel: Effects of elongation on the phase behavior of the Gay-Berne fluid, *Physical Review E*, **57**:6 (1998), 6685–6699.

[51] R. Memmer: Liquid crystal phases of achiral banana-shaped molecules: A computer simulation study, *Liquid Crystals*, **29**:4 (2002), 483–496.

[52] S. J. Johnston, R. J. Low, & M. P. Neal: Computer simulation of apolar bent-core and rodlike molecules, *Physical Review E*, **65**:5 (2002), 051706.

[53] E. Lääräkä: *Data-analyysin perusmenetelmät*, Department of Mathematical Sciences, University of Oulu, 2015.

[54] A. R. Leach: *Molecular modelling: Principles and Applications*, Second Edition, Pearson Education Ltd., 2001.

[55] I. Newton: *Philosophiae Naturalis Principia Mathematica*, J. Societatis Regiae ac Typis J. Streater, 1687.

[56] J. Tuorila: *763620S Statistical Physics*, University of Oulu, 2012.

[57] А. А. Марков: Распространение закона больших чисел на величины, зависящие друг от друга, *Известия Физико-математического общества при Казанском университете*, **15**: (1906), 135—156.

[58] L. Boltzmann: *Vorlesungen über Gastheorie, II. Theil: Theorie van der Waals'; Gase mit zusammengesetzten Molekülen; Gasdissociation; Schlussbemerkungen*, Vorlesungen über Gastheorie, J.A. Barth, 1898.

[59] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, & E. Teller: Equation of state calculations by fast computing machines, *The Journal of Chemical Physics*, **21**:6 (1953), 1087–1092.

[60] D. Frenkel & B. Smit: *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, 2002.

[61] I. McDonald: NpT-ensemble Monte Carlo calculations for binary liquid mixtures, *Molecular Physics*, **23**:1 (1972), 41–58.

[62] E. de Miguel, E. Martin del Rio, & F. J. Blas: Stability of smectic phases in the Gay–Berne model, *The Journal of Chemical Physics*, **121**:22 (2004), 11183–11194.

[63] L. Verlet: Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules, *Physical Review*, **159**:1 (1967), 98–103.

[64] H. Flyvbjerg & H. G. Petersen: Error estimates on averages of correlated data, *The Journal of Chemical Physics*, **91**:1 (1989), 461–466.

[65] B. Efron: Computers and the theory of statistics: Thinking the unthinkable, *SIAM Review*, **21**:4 (1979), 460–480.

[66] E. Kreyszig: *Advanced Engineering Mathematics*, Tenth edition, John Wiley & Sons, Inc., 2011.

[67] R. Ahlrichs, M. Bär, M. Häser, H. Horn, & C. Kölmel: Electronic structure calculations on workstation computers: The program system turbomole, *Chemical Physics Letters*, **162**:3 (1989), 165–169.

[68] J. Lintuvuori, M. Straka, & J. Vaara: Nuclear magnetic resonance parameters of atomic xenon dissolved in Gay–Berne model liquid crystal, *Physical Review E*, **75**:3 (2007), 031707.

[69] G. Luckhurst & P. Simmonds: Computer simulation studies of anisotropic systems, *Molecular Physics*, **80**:2 (1993), 233–252.

[70] *Nonlinear least-squares method documentation*, The SciPy community, 2018, url: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html` (visited on 05/05/2018).

[71] R. Byrd, R. Schnabel, & G. Shultz: A trust region algorithm for nonlinearly constrained optimization, *SIAM Journal on Numerical Analysis*, **24**:5 (1987), 1152–1170.

[72] G. Pelzl, S. Diele, & W. Weissflog: Banana-shaped compounds — A new field of liquid crystals, *Advanced Materials*, **11**:9 (1999), 707–724.

[73] T. Niori, T. Sekine, J. Watanabe, T. Furukawa, & H. Takezoe: Distinct ferroelectric smectic liquid crystals consisting of banana shaped achiral molecules, *Journal of Materials Chemistry*, **6**:7 (1996), 1231–1233.

[74] M. A. Bates & G. R. Luckhurst: Biaxial nematic phases and V-shaped molecules: A Monte Carlo simulation study, *Physical Review E*, **72**:5 (2005), 051702.

[75] A. Chrzanowska: Mean-field model of boomerang nematic liquid crystals with diminished coupling of molecular uniaxial and biaxial susceptibilities, *Physical Review E*, **99**:6 (2019), 062703.

[76] M. A. Bates: Influence of flexibility on the biaxial nematic phase of bent core liquid crystals: A Monte Carlo simulation study, *Physical Review E*, **74**:6 (2006), 061702.